DESCENDING CONTROL IN THE AUDITORY SYSTEM

EDITED BY: David Pérez-González, Miguel A. Merchán, David K. Ryugo and Erika Skoe

PUBLISHED IN: Frontiers in Neuroscience and Frontiers in Systems Neuroscience

frontiers Research Topics

frontiers

Frontiers eBook Copyright Statement

The copyright in the text of individual articles in this eBook is the property of their respective authors or their respective institutions or funders. The copyright in graphics and images within each article may be subject to copyright of other parties. In both cases this is subject to a license granted to Frontiers.

The compilation of articles constituting this eBook is the property of Frontiers.

Each article within this eBook, and the eBook itself, are published under the most recent version of the Creative Commons CC-BY licence. The version current at the date of publication of this eBook is CC-BY 4.0. If the CC-BY licence is updated, the licence granted by Frontiers is automatically updated to the new version.

When exercising any right under the CC-BY licence, Frontiers must be attributed as the original publisher of the article or eBook, as applicable.

Authors have the responsibility of ensuring that any graphics or other materials which are the property of others may be included in the CC-BY licence, but this should be checked before relying on the CC-BY licence to reproduce those materials. Any copyright notices relating to those materials must be complied with.

Copyright and source acknowledgement notices may not be removed and must be displayed in any copy, derivative work or partial copy which includes the elements in question.

All copyright, and all rights therein, are protected by national and international copyright laws. The above represents a summary only. For further information please read Frontiers' Conditions for Website Use and Copyright Statement, and the applicable CC-BY licence.

ISSN 1664-8714
ISBN 978-2-88976-218-7
DOI 10.3389/978-2-88976-218-7

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

DESCENDING CONTROL IN THE AUDITORY SYSTEM

Topic Editors:
David Pérez-González, University of Salamanca, Spain
Miguel A. Merchán, University of Salamanca, Spain
David K. Ryugo, Garvan Institute of Medical Research, Australia
Erika Skoe, University of Connecticut, United States

Citation: Pérez-González, D., Merchán, M. A., Ryugo, D. K., Skoe, E., eds. (2022). Descending Control in the Auditory System. Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-88976-218-7

Table of Contents

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 21 January 2021
doi: 10.3389/fnins.2020.625721

Deafness Weakens Interareal Couplings in the Auditory Cortex

*Prasandhya Astagiri Yusuf[1,2,3], Peter Hubka[2,3], Jochen Tillein[2,3,4,5], Martin Vinck[6,7] and Andrej Kral[2,3,8]**

[1] Department of Medical Physics/Medical Technology Core Cluster IMERI, Faculty of Medicine, University of Indonesia, Jakarta, Indonesia, [2] Institute of AudioNeuroTechnology, Hannover Medical School, Hanover, Germany, [3] Department of Experimental Otology of the ENT Clinics, Hannover Medical School, Hanover, Germany, [4] Department of Otorhinolaryngology, Goethe University, Frankfurt am Main, Germany, [5] MedEL Company, Innsbruck, Austria, [6] Ernst Strüngmann Institut for Neuroscience in Cooperation with Max Planck Society, Frankfurt, Germany, [7] Donders Centre for Neuroscience, Radboud University, Department of Neuroinformatics, Nijmegen, Netherlands, [8] Department of Biomedical Sciences, School of Medicine and Health Sciences, Macquarie University, Sydney, NSW, Australia

The function of the cerebral cortex essentially depends on the ability to form functional assemblies across different cortical areas serving different functions. Here we investigated how developmental hearing experience affects functional and effective interareal connectivity in the auditory cortex in an animal model with years-long and complete auditory deprivation (deafness) from birth, the congenitally deaf cat (CDC). Using intracortical multielectrode arrays, neuronal activity of adult hearing controls and CDCs was registered in the primary auditory cortex and the secondary posterior auditory field (PAF). Ongoing activity as well as responses to acoustic stimulation (in adult hearing controls) and electric stimulation applied via cochlear implants (in adult hearing controls and CDCs) were analyzed. As functional connectivity measures pairwise phase consistency and Granger causality were used. While the number of coupled sites was nearly identical between controls and CDCs, a reduced coupling strength between the primary and the higher order field was found in CDCs under auditory stimulation. Such stimulus-related decoupling was particularly pronounced in the alpha band and in top–down direction. Ongoing connectivity did not show such a decoupling. These findings suggest that developmental experience is essential for functional interareal interactions during sensory processing. The outcomes demonstrate that corticocortical couplings, particularly top-down connectivity, are compromised following congenital sensory deprivation.

Keywords: congenital deafness, predictive coding, bottom-up, top–down, cochlear implant, synchronization

OPEN ACCESS

Edited by:
David K. Ryugo, Garvan Institute of Medical Research, Australia

Reviewed by:
David R. Moore, Cincinnati Children's Hospital Medical Center, United States
Benjamin D. Auerbach, University at Buffalo, United States
Dexter Irvine, Monash University, Australia

***Correspondence:**
Andrej Kral
kral.andrej@mh-hannover.de

Specialty section:
This article was submitted to Auditory Cognitive Neuroscience, a section of the journal Frontiers in Neuroscience

Received: *03 November 2020*
Accepted: *30 December 2020*
Published: *21 January 2021*

Citation:
Yusuf PA, Hubka P, Tillein J, Vinck M and Kral A (2021) Deafness Weakens Interareal Couplings in the Auditory Cortex. Front. Neurosci. 14:625721. doi: 10.3389/fnins.2020.625721

INTRODUCTION

The auditory cortex is composed of a number of cortical areas with different functional roles (Malhotra et al., 2004; Winer and Lee, 2007). Together, these areas form a functional unit that allows constructing and perceiving sensory objects (Kral and Sharma, 2012; Bizley and Cohen, 2013). Only limited information exists on how these areas interact during such processes (Valentine and Eggermont, 2001), and it remains unclear how this interaction develops after birth. While it

has been demonstrated that developmental hearing experience shapes the functional properties of individual brain areas (e.g., Klinke et al., 1999; Chang and Merzenich, 2003; Fallon et al., 2009), the role of experience for integration of cortical areas into a functionally unified auditory cortex is unclear. Despite a lot of effort in investigation of brain connectome (defined as the totality of all connections of the brain), only rudimentary information exists on its developmental constraints.

The question of developmental auditory experience is of particular relevance given that cochlear implants (CIs) restore hearing in congenitally deaf children (Kral and O'Donoghue, 2010). Developmental absence of hearing is accompanied by severe deficits in stimulus feature perception if hearing is restored late in life (Busby and Clark, 1999; Wei et al., 2007). On the other hand, CIs can compensate the deficits and provide access to spoken language with remarkable outcomes if implantations are performed within an early critical period (Manrique et al., 1999; Ponton and Eggermont, 2001; Sharma et al., 2002, 2005; Niparko et al., 2010). Later implantations are typically not successful because, in addition to the loss of (high) juvenile plasticity, congenital deafness strongly interferes with cortical development (review in Kral and Sharma, 2012; Kral et al., 2019): it (i) leads to delays in functional synaptogenesis and augmentation of functional synaptic 'pruning,' (ii) reduces the computational power of cortical networks and (iii) yields abnormally functioning cortical microcircuits. Furthermore, some cortical areas undergo a cross-modal reorganization (Rauschecker, 1995).

Cats have 2 primary and 11 higher-order auditory cortical areas (Rouiller et al., 1991; Winer and Lee, 2007). The posterior auditory field (PAF) is one of the secondary auditory fields (Stecker et al., 2003; Lee and Middlebrooks, 2013). It is part of the "where" pathway as defined in cats, primates and humans (Rauschecker and Tian, 2000; Lomber and Malhotra, 2008). Both bottom–up (e.g., from primary field A1 to secondary field PAF) and top-down (e.g., from PAF to A1) information flow are involved in its function (review in Hackett, 2011). The absence of hearing from birth leads to cross-modal reorganization of PAF, which becomes responsible for supranormal peripheral visual localization in congenitally deaf cats (CDCs) (Lomber et al., 2010). Primary field A1, on the other hand, is not involved in visual or somatosensory reorganizations (Kral et al., 2003; Lomber et al., 2010). A plausible hypothesis is therefore that A1 and PAF show an interareal decoupling in congenital deafness (Kral and Sharma, 2012). Here we test this hypothesis.

Several measures of connectivity have been described (Friston, 2011; Avena-Koenigsberger et al., 2018):

1. Structural connectivity is provided by the anatomical presence of connections (fiber tracts) between the structures of interest. Structural connectivity is typically analyzed by tracer studies in animals or diffusor tensor imaging in humans.
2. Functional connectivity defines statistical dependence among remote physiological events, as frequently analyzed using amplitude correlations or phase coherence, the latter being less dependent on individual response properties. Effective connectivity defines the influence one neural system has on another, either at synaptic or at population level, and is directional. Directional measures such as Granger causality (GC) are used to quantify the effective connectivity.

Structural connectivity provides a scaffold for functional connectivity, but structural and functional connectivity correlate only weakly (Suárez et al., 2020) since functional connectivity additionally captures the dynamics of interactions over time, and involves synaptic efficacy and responsiveness of target structures to patterns stored in the network (Avena-Koenigsberger et al., 2018). Furthermore, functional connectivity may result from common inputs that direct structural connections do not reveal but are functionally relevant for processing (Suárez et al., 2020).

Since the structural connectivity between A1 and PAF is generally preserved in both directions in CDCs (Barone et al., 2013; Butler et al., 2017), the aim of the present study was to compare functional and effective connectivity between A1 and PAF in hearing and deaf cats.

An efficient way to quantify functional connectivity is using the proxy of synchronization of band-specific neuronal activity (Fries, 2005; Womelsdorf et al., 2007; Buzsáki, 2009). Local field potentials allow such analysis (Fontolan et al., 2014; Kornblith et al., 2016). In auditory and visual system, increased synchronization of activity in theta and gamma bands contributes to bottom–up interareal influences, while the increase in alpha and beta bands contribute to top–down influence (Fontolan et al., 2014; van Kerkoerle et al., 2014; Bastos et al., 2015; Michalareas et al., 2016). The influence of congenital deafness on such synchronization is unknown.

As a higher-mammal model of complete sensory deprivation, congenitally deaf (white) cats (CDCs) were used here (Kral and Lomber, 2015). The organization of the auditory cortex in CDCs has been defined functionally and anatomically, including detailed functional maps of fields A1 and the anatomically surrounding fields (e.g., Kral et al., 2006, 2009; Berger et al., 2017). Auditory responses in PAF of CDCs have been characterized previously, too (Yusuf et al., 2017). The present study takes advantage of these previous observations.

In the present study, we compare invasive cortical recordings with multielectrode arrays in three groups of animals: adult hearing cats stimulated acoustically (acoustic controls, ACs), adult CDCs stimulated electrically with CIs, and adult hearing cats likewise stimulated with CIs (electric controls, ECs) following acute destruction of hair cells to prevent electrophonic responses (Sato et al., 2016). These results in two possible comparisons: (i) Whereas CDC and EC receive the same stimulus, they differ in their developmental sensory experience. (ii) AC and EC differ in the stimulus but have the same developmental sensory experience and thus a "similar brain." This latter comparison thus provides information on the influence of stimulus modality (acoustic vs. unknown electric) on the stimulus response.

Phase coherence measures and GC were used to quantify the connectivity strength and the directionality of A1 – PAF interaction in response to auditory stimulation. Phase coherence

is independent of response power (amplitude). Using these connectivity measures we tested the hypotheses whether the artificial electric stimulus generates less interareal interaction than the known acoustic stimulus, and whether CDCs show fundamentally reduced interareal interaction as a consequence of the total absence of hearing.

MATERIALS AND METHODS

Subjects

Fifteen cats, ten adult hearing cats (hearing controls) and five adult CDCs were used in the present study. The details of the experimental procedure were described in previous publications (e.g., Yusuf et al., 2017) and will be briefly recapitulated here. The CDCs were selected from a colony of deaf white cats on the basis of absence of auditory brainstem responses at 120 dB SPL in a hearing screening after birth (Heid et al., 1998; Kral and Lomber, 2015). Each animal's hearing status was confirmed at the beginning of the acute experiments in all animals (for details, see e.g., Berger et al., 2017).

To activate the auditory system in CDCs, the auditory nerve was stimulated electrically using a custom-made CI. As a control for the deaf group, hearing animals were acutely deafened prior to cochlear implantation (intracochlear neomycin application) to prevent responses from healthy hair cells (known as electrophonic hearing, Sato et al., 2016, 2017). Additionally, acoustically stimulated hearing animals were included so that the natural connectivity elicited by acoustic stimuli could be investigated. Thus, the study included three animal groups: ACs (n = 6), ECs (n = 6), and CDCs (n = 5). Of the hearing animals, two were first stimulated acoustically and subsequently stimulated electrically in order to confirm, at the individual level, the effects observed in the group data. Consequently, while 6 animals were in both control groups, only 10 hearing cats were used in total.

The experiments were approved by the local state authorities and were performed in compliance with the Guidelines of the European Community for the care and use of laboratory animals (EUVD 86/609/EEC) and the German Animal Welfare Act (TierSchG).

Experimental Procedures

All animals were premedicated with 0.25 mg atropine i.p. and initially anesthetized with ketamine hydrochloride (24.5 mg/kg, Ketavet, Parker-Davis, Germany) and propionyl promazine phosphate (2.1 mg/kg, Combelen, Bayer, Germany). They were then tracheotomized and artificially ventilated with 50% O_2 and 50% N_2O, with the addition of 0.2–1.5% concentration of isoflurane (Lilly, Germany) to maintain a controlled depth of anesthesia in desynchronized cortical state identified by suppression index values within between 1 and 3, by absence of burst-suppression periods and absence of spindles/bursting (Land et al., 2012). End-tidal CO_2 was continuously monitored and maintained at 4%, and the core temperature was kept at 37.5 – 38.0°C using a homeothermic blanket connected to a rectal temperature probe. Monitoring of the animal's status also involved blood gas concentration measurements, pH, bicarbonate concentration and base excess, glycemia and oxygen saturation determined in capillary blood. A modified Ringer's solution containing bicarbonate and plasma expander was infused i.v. through a venous catheter to supply volume with additional bicarbonate depending on the acid-base status. Use of a higher mammal allows guaranteeing a constant (stable) overall condition of the animal by monitoring and correction of the acid-base balance performed every 12 h throughout the experiments (48–72 h). Furthermore, continuous monitoring of the electrocardiogram, electroencephalogram, breathing pressure and capnometry ensured optimal vital state throughout the whole experiment.

Following tracheotomy, placement of venous and urine catheter, and removal of both pinnae in order to directly access the tympanic membrane for closed-system acoustic stimulation, the animal's head was fixed in a stereotactic frame (Horsley-Clarke). Both bullae and ear canals were subsequently exposed. To record auditory brainstem responses (ABRs), a small trephination was drilled at the vertex of the skull and a silver-ball electrode (diameter 1 mm) was attached epidurally. The indifferent electrode used for the recordings was inserted medially into the neck muscles.

Hearing status was verified using ABRs with 50 μs condensation clicks applied through a closed system directly to the tympanic membrane using a calibrated speaker (DT48, Bayer Dynamics, Germany) at levels up to 120 dB SPL. Brainstem evoked signals were recorded using an epidural vertex electrode against a reference at the midline of the neck, were preamplified (60 dB, Otoconsult V2 low-impedance amplifier), amplified at a second stage (40 dB, Otoconsult Amplifier-Filter F1, filters 0.010–10 kHz) and recorded using National Instruments MIO cards (National Instruments, Munich, Germany). The signals were averaged (200 sweeps, repetition rate 33 Hz, Audiology Lab, Otoconsult, Frankfurt am Main, Germany). Absence of acoustically evoked brainstem responses (including wave I, generated within the auditory nerve) to clicks above 120 dB SPL verified complete deafness. In hearing cats, the thresholds were less than 40 dB SPL before the animals were deafened by slow instillation of 300 μl of neomycin sulfate into the scala tympani (within 5 min.). The Neomycin was left in place for a further 5 min. and subsequently washed out by slow instillation of Ringer's solution. Total absence of brainstem evoked responses verified that the deafening procedure was successful. For electrical stimulation, hearing cats and CDCs were implanted with a CI inserted via the round window. The implant consisted of a medical-grade silicone tube with five intrascalar contacts: a small golden sphere at the tip (diameter 0.8 mm) and four golden rings, the distance between all electrodes being 1 mm. The intrascalar part of the implant was tapered in the apical direction from a diameter of 1.6 mm to 0.8 mm. The extracochlear silicone tube had a diameter of 1.6 mm. The gold contacts were connected to a seven-strand Teflon-coated stainless-steel braided wire. The stimulation mode was wide bipolar (most apical vs. the fourth intracochlear electrode in the basal direction; distance between active electrodes was thus 3 mm).

Electrically evoked auditory brainstem response (E-ABR) to single biphasic pulses was recorded and the lowest current levels evoking a brainstem response (E-ABR-threshold currents) were determined. For this purpose, charge-balanced biphasic pulses (200 μs/phase, repetition rate 33 pps) were applied to the CI using wide bipolar stimulation (most apical and most basal electrode). Stimulation was performed with optically isolated current sources (CS1, Otoconsult, Frankfurt am Main, Germany).

Stimulation and Recording

Trephination was performed above the auditory cortex and the dura was removed. The cortex was photographed to document the recording positions. Using an ORIEL motorized x-y-z micromanipulator (1 μm precision in all directions), a silver-ball macroelectrode (diameter 1 mm) was positioned at a regular raster of nine cortical positions on the primary auditory cortex (field A1). The dorsal end of the posterior ectosylvian sulcus was used as a reference point. Signals (local field potentials, LFPs) recorded in response to an electric biphasic pulse applied through a CI were preamplified (60 dB, Otoconsult V2 low-impedance amplifier), amplified at a second stage (20 dB, Otoconsult Amplifier-Filter F1, filters 0.010–10 kHz), recorded using MIO cards and averaged (100 sweeps, repetition rate 1.97 Hz). The signals were stored and threshold current levels were evaluated at all recording positions with a precision of ±1 dB.

In order to determine the extent of the cortical activated region, a Ringer-filled glass microelectrode (impedance < 6 MΩ) was used for mapping the field A1. LFPs on the cortical surface were recorded at 75–150 cortical positions during stimulation with the CI, using single biphasic pulses (200 μs/phase, wide bipolar stimulation at both the ipsilateral and contralateral ear, stimulation current 10 dB above the lowest cortical threshold determined using the macroelectrode). The stimuli were applied at a repetition rate of ∼2 pps. Recorded signals were bandpass filtered (10–9000 Hz) and amplified 5000 times (Neuralynx Cheetah, Bozeman, MT, United States). The data were digitized using a NI PCIe 6259 MIO card at a sampling rate of 25 kHz per channel. Fifty responses were averaged to obtain evoked LFPs. Amplitudes of these middle-latency responses (peak to baseline) were used to construct cortical activation maps and determine the most responsive region in A1, the "hot spots" (Kral et al., 2009).

Simultaneous recordings from the right A1 and PAF were performed contralateral to the stimulated ear. In A1, using a micromanipulator the cortex was penetrated perpendicular to the surface in the 'hot spot' (responses with >300 μV amplitude, Kral et al., 2009) with a single-shank Neuronexus probe (16 contacts, 150 μm spacing, around 1–2 MΩ impedance). The probe was inserted so that the last contact just disappeared into the cortex (penetration depth ∼2400 μm). Since PAF is hidden in a sulcus, the recording electrode could not be inserted radially as in A1. To cover the complete PAF, we recorded the LFP signals from two penetration depths (electrode tip depth at 5,000 and 2,500 μm penetration depth) using a second Neuronexus probe with the same characteristics as the first (**Figure 1**). This was performed through the dorsoventral extent of this field parallel to the course of the posterior ectosylvian sulcus with a penetration-to-penetration distance of ∼500 μm in the dorsoventral direction. All manipulation was performed using micromanipulators (precision ∼1 μm) and under visual control through the operating microscope (OPMI1-H, Zeiss Deutschland, Oberkochen, Germany). Recorded signals were bandpass filtered (1–9000 Hz) and amplified 5,000 times (Neuralynx Cheetah, Bozeman, MT, United States). The data were digitized using a NI PCIe 6259 MIO card at a sampling rate of 25 kHz per channel. During these recordings, the cortex was stabilized by means of a modified Davies chamber (Tillein et al., 2010). The reference for both probes was the vertex silver-ball electrode placed epidurally. Off-line, bipolar derivation of the signals in A1 before connectivity analysis ensured that the reference did not influence connectivity results.

The ACs were stimulated acoustically using three condensation clicks (50 μs duration at 500 pps) at different sound pressure levels. The ECs and CDCs were stimulated using a custom-made CI inserted into the scala tympani through the round window. The stimulus was a train of charge-balanced biphasic pulses (200 μs/phase, repetition rate 500 pps, three pulses in the train applied). Stimulation was in wide bipolar configuration. The acoustic and electric stimuli were applied at a repetition rate of 1/1537 ms, with 30 stimulus repetitions per condition (level). Stimulus increased in 10 dB increments in acoustic stimulation and in 1–2 dB increments in electric stimulation. Stimulus artifacts were removed by linear interpolation of the 6 ms period during stimulation. In a previous study we ensured that this did not introduce any artifacts into frequency-specific signals used (Yusuf et al., 2017).

The stimulation levels for connectivity analysis were chosen according to input–output level functions (levels that reach the saturation of evoked response). Analyses reported in the present study were performed at 40 dB (acoustic) above ABR threshold, while electrical stimulation was administered to ECs and CDCs using three biphasic electric charge-balanced pulses at 6 dB (electric) above the electrically evoked auditory brainstem response (E-ABR) threshold.

Histology

For each animal, at least one penetration for each field was marked by a fluorescent dye (DiI, 1,10-dioctadecyl-3,3,3′,3′-tetramethylindocarbocyanine perchlorate; Invitrogen). Since the probe attachment to the stereotactic frame was constant throughout the experiment, it was possible to extrapolate all penetrations directions from the stained and reconstructed tract. In PAF, histological reconstructions confirmed the correct location within this field in all animals reported.

After the experiments, the animals were transcardially perfused in deep anesthesia. Following thoracotomy, 0.5 ml heparin (Heparin Natrium, Ratiopharm, Ulm, Germany) was injected into the both ventricles. Two liters of 0.9% NaCl solution and two liters of fixative (4% paraformaldehyde) and one liter of 10% sucrose were infused transcardially. The perfusion pressure was kept constant at 120–150 mmHg and monitored using the Perfusion One system (Leica Biosystems, Buffalo Grove, IL, United States). If required, the brain was postfixated in 4% paraformaldehyde and 10% sucrose overnight. For cryoprotection, each brain was placed in 30% sucrose

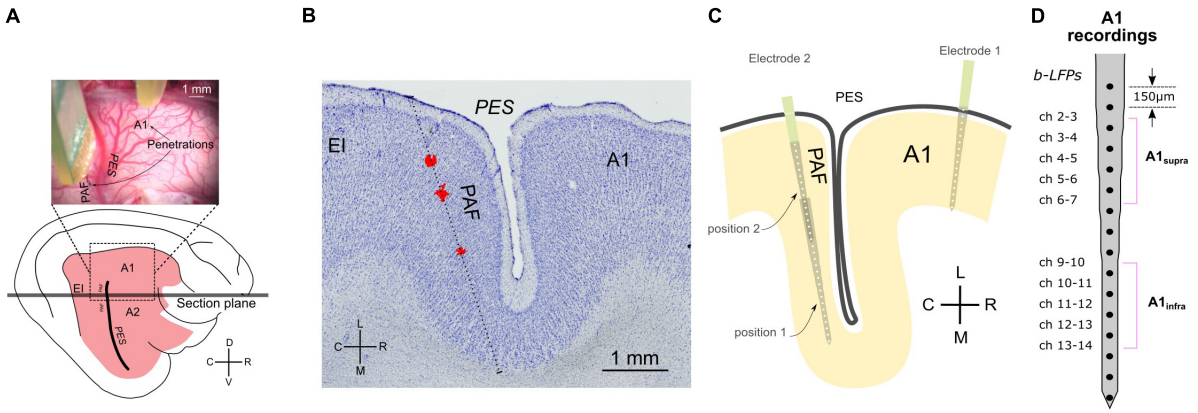

FIGURE 1 | Methodology and recording positions. **(A)** Top: photograph of feline cortex after trephination, revealing the penetration sites in A1 and PAF. Bottom: illustration of entire brain from the same perspective. **(B)** Reconstruction of penetration of the DiI stained probe in a Nissl-stained section. The sectional plane is shown in **(A)**. The stained images were stacked and aligned to reconstruct the penetration. The red deposits shown (DiI) were extracted from several successive florescence images from the same region of the cortex and projected onto the Nissl-stained section. The reconstructed direction of penetration is shown as a dotted line. **(C)** Schematic illustration of electrode penetrations in A1 and PAF. In PAF, dense mapping allowed capture of auditory responses in each animal. Using two recording depths, each penetration includes 32 recording sites in total. **(D)** Channels of bipolar derivation LFP (b-LFP) in A1 recordings, grouped into supragranular ($A1_{supra}$) and infragranular ($A1_{infra}$) layers. A1, primary auditory cortex; EI, intermediate area of the posterior ectosylvian gyrus; PAF, posterior auditory field; PES, posterior ectosylvian sulcus; V, ventral; D, dorsal; R, rostral; C, caudal; L, lateral; M, medial.

solution until it sank. Subsequently, the brain was blocked, frozen at −80°C and cut at −20°C using a Leica Cryostat CM3050S (Leica Microsystems GmbH, Wetzlar, Germany) in section 50 µm thick. The sections were first photographed to reveal the DiI in fluorescent mode using a Keyence BZ-9000 microscope and subsequently stained using Nissl staining and SMI-32. For reconstruction, native fluorescence images were combined with the same Nissl-stained sections.

Layers in A1 were grouped into supragranular, granular and infragranular based on the reconstructions of penetrations. The Nissl staining reveals the border of layer IV to layer V (Berger et al., 2017). Additionally, current source density measures (CSDs) that show a typical sequence of middle source in layer III and deep sink in layer V, with an initial sink followed by a source in layer IV between them (Kral et al., 2006), confirm this differentiation.

Time Domain Analysis

All data processing and analyses mentioned in this section were performed offline using the FieldTrip toolbox[1] (Oostenveld et al., 2011) and custom-made MATLAB scripts (Mathworks Inc., Aachen, Germany). Occasional noisy recordings caused by unstable probe contacts, channels with artifacts and occasional trials with spindles were not included in the analyses.

Discrete Fourier transformation (DFT) filters at 50 and 100 Hz were applied to remove power line artifacts. The detrend (demean) procedure was applied to the LFP signals to remove any possible DC shift in the recordings. We reduced the far-field components in A1 by subtracting every two adjacent channels within an electrode shank from each other, yielding the bipolar derivation LFP (b-LFP) signals. We removed the transient evoked components by subtracting the time domain averaged signal from each trial, allowing the analysis of the non-phase-locked part only (Donner and Siegel, 2011; Siegel et al., 2012). In the following, all connectivity analyses were computed from the non-phase-locked signals of bipolar derivation LFPs in A1 and non-phase-locked unipolar LFPs in PAF.

Spectrum Analyses

Hanning-tapered Fourier transformation was computed based on the LFP data in the prestimulus/baseline time window (−400 to −1 ms) and in the late-latency poststimulus time window (200–600 ms). Frequencies from 1 to 128 Hz with 1 Hz linear increments were subsequently analyzed. Power spectra were generated by taking the absolute square of the transformation.

Time-frequency representations (TFRs) were computed by means of complex wavelet analysis (using Morlet wavelet, $m = 6$) with 56 logarithmic frequency increments from 4 to 128 Hz, thus capturing the theta (4–8 Hz), alpha (8–16 Hz), beta (16–32 Hz), low-gamma (32–64 Hz), and high-gamma (64–128 Hz) frequency bands, in an equal number of bins (Hipp et al., 2011).

Functional Connectivity

We computed the phase coherence between A1 and PAF electrodes using debiased weighted phase-lag index (Vinck et al., 2011) (WPLId) and pairwise phase consistency (Vinck et al., 2010) (PPC). These methods are insensitive to sample size bias (WPLId) or unbiased to sample size (PPC), which fits with the availability of 30 trials in this study. The values range from zero (negative values due to limited sampling were corrected to zero) to one (maximum coherence).

As WPLId includes only the imaginary part of the cross-spectrum, it is sensitive only to the true interaction between two signals but not to the common reference and far-field (volume

[1]http://www.ru.nl/fcdonders/fieldtrip/

conduction) signals (Vinck et al., 2011). A higher signal-to-noise ratio is also found in comparison with other connectivity measures based on the imaginary component of the cross-spectrum (Phillips et al., 2014; Babapoor-Farrokhran et al., 2017). Due to its sensitivity in detecting true interaction, here WPLId was used for defining *significant coupling*. The WPLId value was *z*-score normalized to its standard deviation (Nolte et al., 2008), estimated by the applying leave-one-out jackknife procedure (Richter et al., 2015) from the multiple observations (trials), as follows

$$wPLId_z = \frac{wPLId}{std(wPLId)} \quad (1)$$

This enables phase coherence to be reliably indexed using z-scores. The *significantly coupled channel pairs* were computed by thresholding the couplings with maximum *z*-score values exceed the equivalent of $p < 0.05$ (Bonferroni corrected). Subsequently, we recomputed the functional connectivity using the PPC method, only including channel-pairs with significant coupling. PPC yields results proportional to true angle distribution and therefore we focused on this method (WPLId results are available and were consistent in outcome with PPC).

The PPC method computes the vector dot product (i.e., the projection of one vector onto another) for all given trial pairs of relative phases. The higher the phase consistency across trials, the smaller the angular distance, and hence the higher the dot products for each pair. The PPC value is defined as the average of the dot product across all available pairs [$0.5 * N * (N\text{-}1)$, where N denotes number of trials] (Vinck et al., 2010). Unless specifically mentioned, all PPC values are presented in change to baseline, subtracting the late-latency poststimulus time PPC with the prestimulus time PPC.

Effective Connectivity

Effective connectivity was computed using the non-parametric GC (Dhamala et al., 2008). GC analysis is useful for quantifying bidirectional interaction, i.e., separately quantifying GC influence from A1 to PAF ($GC_{A1\rightarrow PAF}$) and the influence from PAF to A1 ($GC_{PAF\rightarrow A1}$). GC spectra were obtained by computing Geweke's frequency domain GC (Geweke, 1982) and the spectral factorization technique was used for complex cross-spectral density, obtained from the Fourier transformation. Non-parametric GC is advantageous since it does not require model order for autoregressive computation (as in the parametric GC), but has a drawback: cross-spectral density yields a smoothened shape (Bastos and Schoffelen, 2016). GC values are presented as change to baseline, subtracting the late-latency poststimulus time GC with the prestimulus time GC.

Directionality (GC_{flow}) was computed as $GC_{A1\rightarrow PAF}$ minus $GC_{PAF\rightarrow A1}$. Consequently, positive values represent the domination of bottom-up interaction (A1→PAF) while negative values represent the domination of top–down interaction (PAF→A1).

We computed reversed-time GC to check for any false GC analysis results due to the presence of correlated and uncorrelated noise in the signal (Vinck et al., 2015). Time reversal of the signal prior to GC computation should consequently reverse the domination of directionality (GC_{flow}). The presence of noise in the signal will not change this flow domination, i.e., from a positive to a negative, or from a negative to a positive value (Vinck et al., 2015). Therefore, time-reversing the signal is an effective procedure for confirming the directionality from GC analysis. We excluded channel pairs from the grand average computation where the requirement for 'flipped directionality' in the reversed-time GC was not satisfied.

Statistics

We compared acoustic and ECs to reveal the influence of stimulation mode, and ECs with CDCs to reveal the effect of congenital sensory deprivation. ACs could not be directly compared with CDCs due to several biasing factors: they differed not only in developmental sensory experience but also in the mode of stimulation (acoustic vs. electric) and in the presence of hair cells generating spontaneous activity. Thus, differences would be equivocal with respect to several factors.

The differences between each pair of groups (CDCs vs. ECs and acoustic vs. electric controls) for the spectrum-based analyses were tested using the Wilcoxon rank-sum test, corrected with false discovery rate procedure (Benjamini and Yekutieli, 2001). For the TFR-based analyses, we used non-parametric cluster-based permutation statistics (Maris and Oostenveld, 2007) with 1,000 random permutations under the null hypothesis (cluster α threshold 0.5%, two-tail significant α value = 0.25%) – (i) compared against zeros for significant increase and decrease in each site-pair and group and (ii) compared between groups yielding pair comparison in each site-pair.

TABLE 1 | Cortical depths for each electrode of the probe in A1 over the range of deviations between 0° and 14° from perpendicular as observed in the present experiments.

	Layer border [μm]	Unipolar		Bipolar	
		Channel #	Cortical depth [μm]	Channel #	Cortical depth [μm]
Supragranular	150–900	2 to 7	∼146–900	2–3 to 6–7	218–825
Granular	900–1150	8	1019–1050	7–8 to 8–9	946–1125
Infragranular	>1150	9 to 16	>1164	>9–10	>1237

In total, we analyzed 470 sites in supragranular and 663 sites in infragranular layers of acoustic controls, 787 sites in supragranular and 711 in infragranular layers in electric controls, and 1147 sites in supragranular and 1342 sites in infragranular layers of CDCs.

RESULTS

Local field potentials (LFPs) in primary auditory cortex (A1) and the posterior auditory field (PAF) were recorded in the cortex contralateral to the stimulated ear. The cortex was penetrated perpendicularly to the cortical surface at the most responsive area of A1 (the hot spot, same as in Tillein et al., 2010, 2016; Yusuf et al., 2017) with a multielectrode array. Recordings in PAF were performed throughout the entire dorsoventral extent of the field parallel to the posterior ectosylvian sulcus using another multielectrode array at up to 10 penetrations in PAF of each animal (**Figure 1**). This resulted in layer-specific recordings in A1 and tangential recording tracks in PAF (**Figures 1B,C**). To minimize the contribution of volume conduction effect on connectivity analysis between A1 and PAF, and to localize the sources of LFPs to individual layers, off-line signal subtraction between neighboring channels (bipolar derivation LFP) in A1 was calculated (**Figure 1D**). We determined the cortical depth of each channel and grouped them to the corresponding layers within A1 (**Table 1**, see Berger et al., 2017). In the following, we combined A1 recordings within supragranular layers and within infragranular layers (denoted as $A1_{supra}$ and $A1_{infra}$). Layer IV in A1 was excluded from the subsequent statistical analysis because long-range corticocortical connections are not present in layer IV of A1: its inputs originate in the thalamus (Mitani and Shimokouchi, 1985; Markov et al., 2014). In PAF, due to the tangential course of penetration, precise identification of recorded layers was not possible for all electrode contacts. The use of unipolar signals allowed additionally increasing the sensitivity for coupling by capturing

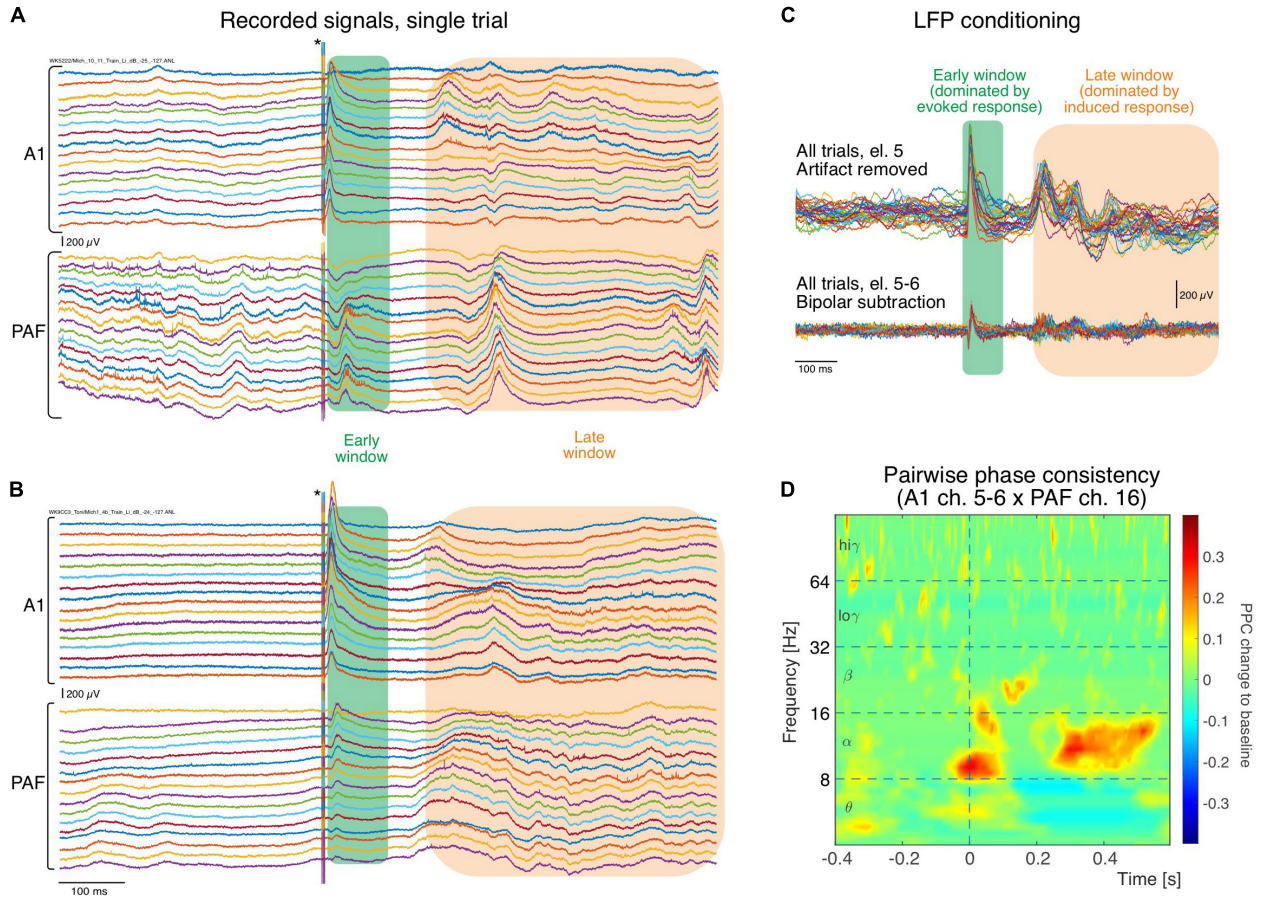

FIGURE 2 | Single raw trace LFP examples recorded simultaneously in different cortical positions (in both fields arranged from surface to deep) during stimulation. **(A)** In an electric control, the electric stimulus generates a large artifact that is discernible in all recordings that lasts throughout the 6 ms of stimulus duration (asterisk). Following the stimulus a short latency response (green rectangle) is observed in all traces of recordings, larger in A1 and smaller in PAF. Two hundred ms after the stimulus, some increased activity can be observed that is less well synchronized to the stimulus than the early response (orange rectangle). **(B)** In the congenitally deaf cats, similar activity in both windows is observed in A1 and PAF (for systematic differences, see Yusuf et al., 2017). **(C)** Example of trial-to-trial variability for one electrode following stimulation artifact removal, obtained from the A1 recording shown in **(A)**. The early window response shows higher trial-to-trial consistency than the late response, corresponding to the previous description of an evoked response caused by thalamic input. In the late window the responses vary between trials in latency/phase and amplitude, typical for induced responses resulting from interaction of activity caused by the stimulus with corticocortical inputs. After bipolar derivation, far-field and common reference influences are eliminated and amplitudes decrease, but early and late responses are preserved. **(D)** Example of pairwise phase consistency computed from recording pair of electrodes 5–6 (bipolar) in A1 and electrode 16 in PAF [the 32nd trace in **(A)**]. There is a strong synchronization of activity in the alpha band and the late window, documenting a stimulus-related coupling of these sites.

signals from sources not directly within the penetration in PAF. Thus for determining the coupling we used local sources in A1 and less local sources in PAF.

An example of the original registered activity (before artifact elimination and bipolar derivation) at 6 dB above threshold in both investigated fields is shown in **Figure 2**. In the individual trials, both in the hearing animal (**Figure 2A**) and in the CDC (**Figure 2B**), fast responses following the stimulus within a time window of <100 ms post stimulus (termed *early window* here) can be observed in both fields, although smaller in amplitude and longer in latency in PAF. Approximately 200 ms after the stimulus, a second increase in activity is observable that is less well synchronized (time-locked) with the onset of the stimulus and thus variable from trial to trial, yet is very different from prestimulus activity (>200 ms termed *late window* here). When a single recording contact is considered, the reduced synchronization relative to stimulus onset in the late window becomes apparent (**Figure 2C**). The synchronized response, predominantly observed in the early window, will be called *evoked response* and the response that is not synchronized, predominantly observed in the late window, will be called *induced response* (for previous detailed analysis, see Yusuf et al., 2017). Using such signals pairwise phase consistency can be computed in a frequency-specific manner (**Figure 2D**). The peak PPC increases after the stimulus reached values of up to 0.4. In some recording positions, peaks in coupling were more pronounced in the early window.

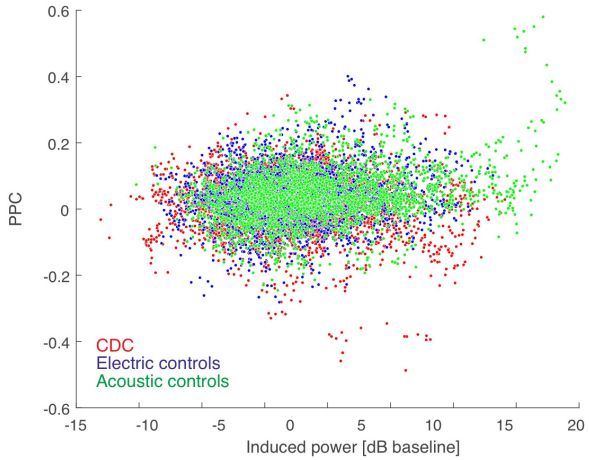

FIGURE 3 | PPC-based connectivity is not a direct consequence of induced power (summed for the pair). Color code shown in the inset. The Spearman correlation coefficient was very low (rho = 0.071 for acoustic controls, rho = 0.017 for electric controls, and rho = 0.002 for CDCs, all $p < 0.05$) and thus induced power contributed minimally to the PPC result. There was no difference between the three groups of animals in power-PPC relation.

Stimulus-Related Connectivity: Congenital Deafness Reduces Top–Down Interactions

We first identified individual simultaneously recorded site-pairs that showed significant couplings at any frequency (*z*-score estimation using jackknife procedure, see section "Materials and Methods"), being around half of all compared electrode pairs in all three groups (ACs: 61%; ECs: 49%; CDCs: 53%). This corresponds to the observation of similar anatomical connectivity between these two cortical areas between deaf and hearing cats (Barone et al., 2013; Butler et al., 2017). Only these coupled site pairs were used for further analysis.

Next we confirmed that our measure of functional connectivity, i.e., PPC, is not dependent on response strength. We compared the PPC as a function of the sum of the power of the induced responses at the two corresponding positions in all coupled pairs (**Figure 3**). The very small correlations show that PPC is not dependent on induced power. This further means that the results of connectivity analysis are (as expected) not the consequence of differences in signal power. Thus, differences in signal power in CDCs compared to ECs, as observed in a previous study (Yusuf et al., 2017), did not determine the coupling results of the present study.

Stimulus-related coupling increases were observed in both the early and late windows in both control groups. Grand mean averages for all three groups investigated are shown in **Figure 4**. The mean values underestimate the PPC increases observed in individual recording pairs (as in **Figure 2D**) due to differences in exact timing and frequency between the pairs, but the grand means reflect the most common features of the couplings and are appropriate for robust statistical comparisons.

We concentrated on the increases in PPC relative to baseline, since these reflect stimulus-related functional coupling between the sites. Both acoustic and electric controls, irrespective of the recorded layer, showed an increase in coupling in beta and gamma bands in the early window (**Figures 4A–D**). Increased coupling was also observed in the alpha band, but this finding was limited to some layer groups only. A second period of alpha coupling appeared in the late window (asterisk, see also **Figure 2**), discernible in both control groups and both layer groups. These coupling increases were significantly different from zero (shown in **Figures 4C,D** as black lines, cluster-based permutation test, two-tail significant $\alpha = 0.25\%$). In both control groups we observed also variable desynchronizations in the late window.

Statistical analysis of the differences between the two control groups is shown by the blue lines in **Figures 4A,B** (cluster-based permutation test, two-tail significant $\alpha = 0.25\%$). In general, the early connectivity as well as the late connectivity (asterisk) were not different. However, smaller "islands" of desynchronization, together with a beta and theta coupling in the late window, were larger in amplitude in ACs. This observation may be related to spontaneous activity from the hearing cochlea.

Consistently, in grand means of both acoustic and electric controls and both layer groups, there was (i) an increase in beta and gamma coupling after the stimulus in the early window and (ii) an increase in alpha coupling in the late window.

In CDCs, only the early couplings were preserved. The early synchronization in beta and gamma band was not different from ECs, but in contrast to ECs the early as well as late alpha synchronization disappeared in CDCs (**Figures 4E,F**; significance

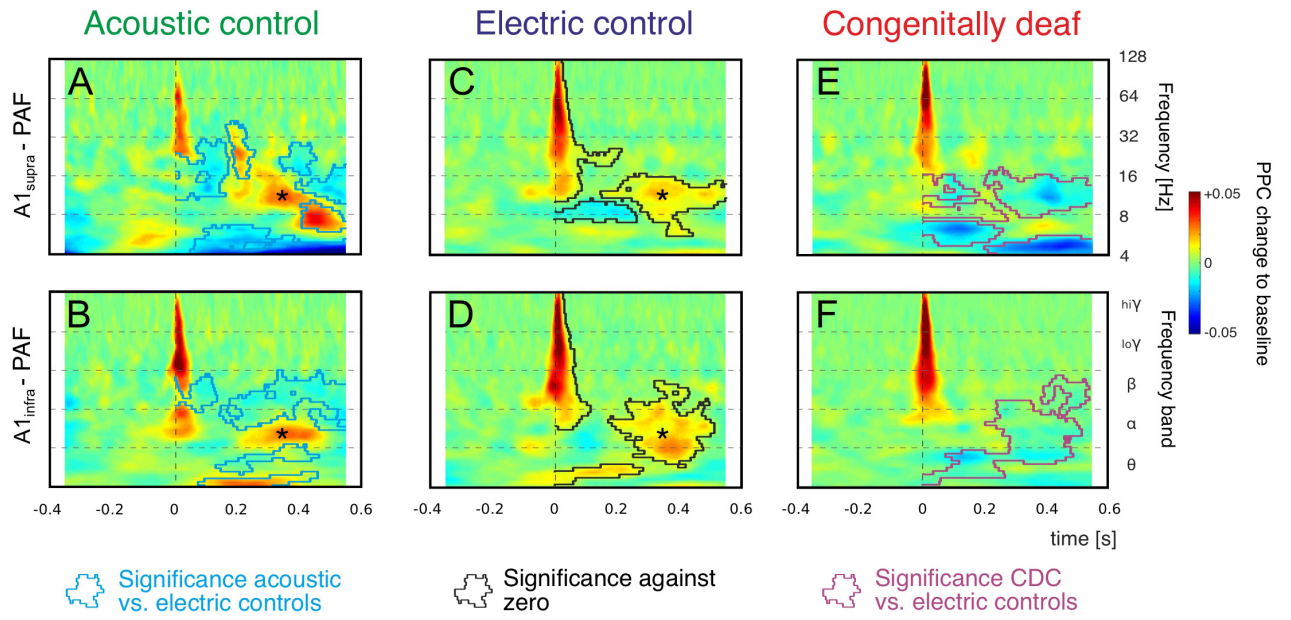

FIGURE 4 | Grand mean averages of time-frequency-based functional connectivity between A1 and PAF computed using pairwise phase consistency (PPC). Functional connectivity computed separately for supragranular layers in A1 ($A1_{supra}$ – PAF, top) and infragranular layers in A1 ($A1_{infra}$ – PAF, bottom) reveal onset responses (near 0 ms) and late responses (>200 ms). Warm (yellow and red) colors represent an increase in synchronization relative to baseline, i.e., a stimulus-related functional connection. **(A,B)**: Acoustic controls: Regions with significant differences to electric controls are outlined with blue lines (non-parametric cluster-based permutation statistical testing, two-tail significant α value = 0.25%). **(C,D)** Electric controls: Regions of significant PPC change to zero are outlined by black lines (comparison to zero, two-tail significant α value = 0.25%), statistical differences to acoustic controls are shown in **(A,B)** as blue lines (non-parametric cluster-based permutation statistical testing, cluster α threshold 0.5%, two-tail significant α value = 0.25%). Alpha band synchronization in the late window (asterisk) is observed in both control groups consistently. **(E,F)** In CDCs, only onset couplings are preserved, all late (>200 ms) couplings disappeared. Regions with significant differences to electric controls are outlined with magenta lines (non-parametric cluster-based permutation statistical testing, two-tail significant α value = 0.25%).

to ECs shown by magenta lines, cluster-based permutation test, two-tail significant α = 0.25%). Thus, CDCs differed greatly from the controls: a part of the early coupling as well as all late coupling between A1 and PAF following an auditory stimulus in the alpha band, consistently found in both controls and all layers, disappeared in CDCs (**Figure 4E**).

The most extensive effect of developmental experience was observed in the late window. In order to further quantify overall effects we pooled the couplings over the entire late poststimulus time window (>200 ms) in the conventional frequency bands. The theta band, while showing mixed effects in controls and only desynchronizations in CDCs (**Figure 4**) has to be treated with caution due to the temporal windows available (prestimulus 400 ms) that are at the limit of the temporal requirements for this band, particularly when relative-to-baseline measures are used. Therefore it was not analyzed further. In gamma bands no effects were observed in the late window. Alpha coupling increased following a sensory stimulus in the late window in acoustic and electric controls, whereas it decreased in CDCs (**Figure 5**; $p = 3.234 * 10^{-17}$ for $A1_{supra}$-PAF and $p = 1.013 * 10^{-9}$ for $A1_{infra}$-PAF, Wilcoxon rank-sum test). Weaker and less consistent effects were observed in the beta band, where particularly ACs showed the alternating periods of synchronization and desynchronization (**Figures 4A,B**), leading to a mean desynchronization if summed over time (**Figures 5A,B**, green bars).

Granger causality was used to determine the directionality of the alpha and beta A1-PAF interactions in the late window, where PPC differences were found. Both A1 to PAF bottom-up interactions and PAF to A1 top–down interactions were quantified. Previous work has shown that alpha (van Kerkoerle et al., 2014; Michalareas et al., 2016) and beta (Bastos et al., 2015; Michalareas et al., 2016; Richter et al., 2018) bands are associated with stimulus-related top–down feedback between sensory areas. In keeping with these findings, auditory stimulation induced a prominent increase (relative to baseline) of alpha- and beta-band top–down GC in ACs (**Figures 6A,B**). Also in ECs the increase in top–down GC was larger than the increase in bottom–up GC (**Figures 6A,B**). Importantly, the CDCs did not show this effect, rather top–down was smaller or same as bottom–up GC, and overall the GC change was small for both supragranular and infragranular layers of A1.

To establish the overall dominant direction of the information flow, we computed the difference between top–down and bottom–up GC, resulting in the Granger flow measure (Fontolan et al., 2014; Babapoor-Farrokhran et al., 2017). Negative Granger flow signifies predominantly top–down-directed interaction, whereas positive Granger flow indicates predominantly bottom–up-directed interaction. The results demonstrate that the stimulus-related change in interaction was a shift toward top–down interaction in both hearing acoustic and electric group (i.e., net top–down interaction). In $A1_{supra}$-PAF coupling,

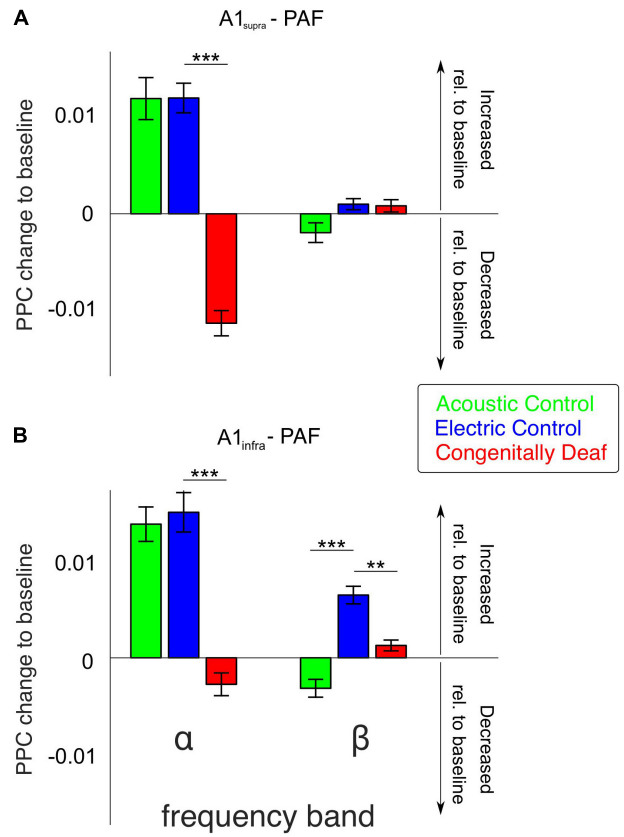

FIGURE 5 | Bar plots of average alpha- and beta-band functional connectivity within the late poststimulus time window for all three groups, separately for $A1_{supra}$ – PAF **(A)** and $A1_{infra}$ – PAF **(B)** couplings. Data shown correspond to the change in PPC relative to baseline (baseline PPC subtracted from poststimulus PPC). Positive (negative) values indicate increased (decreased) functional connectivity relative to pre-stimulus baseline, respectively. Group pairwise comparisons were computed using the two-tailed Wilcoxon rank-sum test. $^{**}P < 0.01$; $^{***}P < 0.001$.

Granger flow in ACs was significantly more negative than in ECs (**Figure 6C**, $p = 1.321 * 10^{-9}$ in the alpha band and $p = 6.684 * 10^{-8}$ in the beta band, Wilcoxon rank-sum test).

Granger causality showed weaker top–down connectivity in ECs compared with ACs in all investigated layers of A1 (**Figures 6C,D**). Additionally, PPC revealed subtle differences in connectivity between acoustic and electric controls (**Figure 5**). We interpret these observations as a consequence of the artificial electrical stimulus highly synchronizing the auditory nerve firing.

Congenitally deaf cats, on the other hand, lost the top-down flow observed in ECs in the alpha band (**Figure 6C**, $p = 4.642 * 10^{-4}$, Wilcoxon rank-sum test). Beta band, where also ECs showed small effects, was not significant (**Figure 6C**, $p = 0.276$, Wilcoxon rank-sum test). Remember, in beta band also the PPC outcomes revealed minimal effects (**Figure 5**). In $A1_{infra}$-PAF coupling, ECs showed stronger net top–down Granger flow in beta-band than CDCs (**Figure 6D**, $p = 0.017$, Wilcoxon rank-sum test) – in-line with the largest beta connectivity found in PPC in $A1_{infra}$-PAF (**Figure 5**).

In total, these findings show that the stimulus mode is affecting the connectivity measures. That observation demonstrates that our methods are sensitive to changes in stimulus properties and that functional connectivity can change if the stimulus changes - even in the brain with same anatomic connectivity and same membrane properties of the cortical neurons involved. Absent developmental hearing experience (electric controls vs. congenitally deaf cats) eliminated the stimulus-related coupling increase in the late time window. The results demonstrate that top–down connectivity is substantially involved in the reduced effective connectivity observed in CDCs.

Ongoing Activity and Connectivity Reveal Layer-Specificity of Deafness Effects

Finally, we tested whether the stimulus-related connectivity could be a mere consequence of resting-state (i.e., ongoing) connectivity. We analyzed ongoing extracellular LFP activities to reveal the power-spectral activity and phase-based connectivity in the absence of an auditory stimulus. It is of importance to emphasize that the groups differed regarding the state of the organ of Corti: whereas ACs had an intact cochlea, in ECs the hair cells were destroyed by intrascalar neomycine injection. Similarly, CDCs did not have surviving hair cells. This is of substantial relevance, because hair cells are the main driver of spontaneous activity in the auditory nerve, providing a tonic drive to the auditory pathway (and auditory cortex; see also discussion). Furthermore, trials containing bursts of activity and spindles were eliminated from the analysis since they may confound connectivity measures (e.g., Valentine and Eggermont, 2001).

The LFP power spectrum revealed a level of ongoing activity significantly higher in ACs than in ECs in almost all frequency bands in $A1_{supra}$, $A1_{infra}$, and PAF (**Figures 7A–C**). Since these two control groups differ with regard to surviving hair cells and hence spontaneous activity in the auditory nerve, this outcome suggests that, up to the level of the secondary auditory cortex, spontaneous activity in the auditory nerve is possibly a significant factor driving cortical ongoing activity.

The CDCs, in a part of the A1 data, exhibited significantly higher ongoing LFP power than ECs in $A1_{supra}$ for the alpha- and beta-band (**Figure 7A**, $p < 0.001$, false discovery rate corrected Wilcoxon rank-sum test). This phenomenon was layer specific: it was not observed in the infragranular layers of A1 (**Figure 7B**, compare anatomical outcomes in Berger et al., 2017). It was also absent in PAF, where the power was in fact significantly lower in CDCs than in ECs in the theta/alpha bands (**Figure 7C**, $p < 0.001$, false discovery rate corrected Wilcoxon rank-sum test).

The ongoing functional connectivity between A1 and PAF, quantified by the pairwise phase consistency (PPC), also revealed layer-specific differences. There was higher baseline phase coherence between $A1_{supra}$ and PAF in ACs compared with ECs which was significant in theta/alpha/beta bands (**Figure 7D**; $p < 0.001$, false discovery rate corrected Wilcoxon rank-sum test). This was consistent with the higher ongoing power in ACs, but remember that PPC is power independent (comp. **Figure 3**). The infragranular layers, despite higher ongoing power

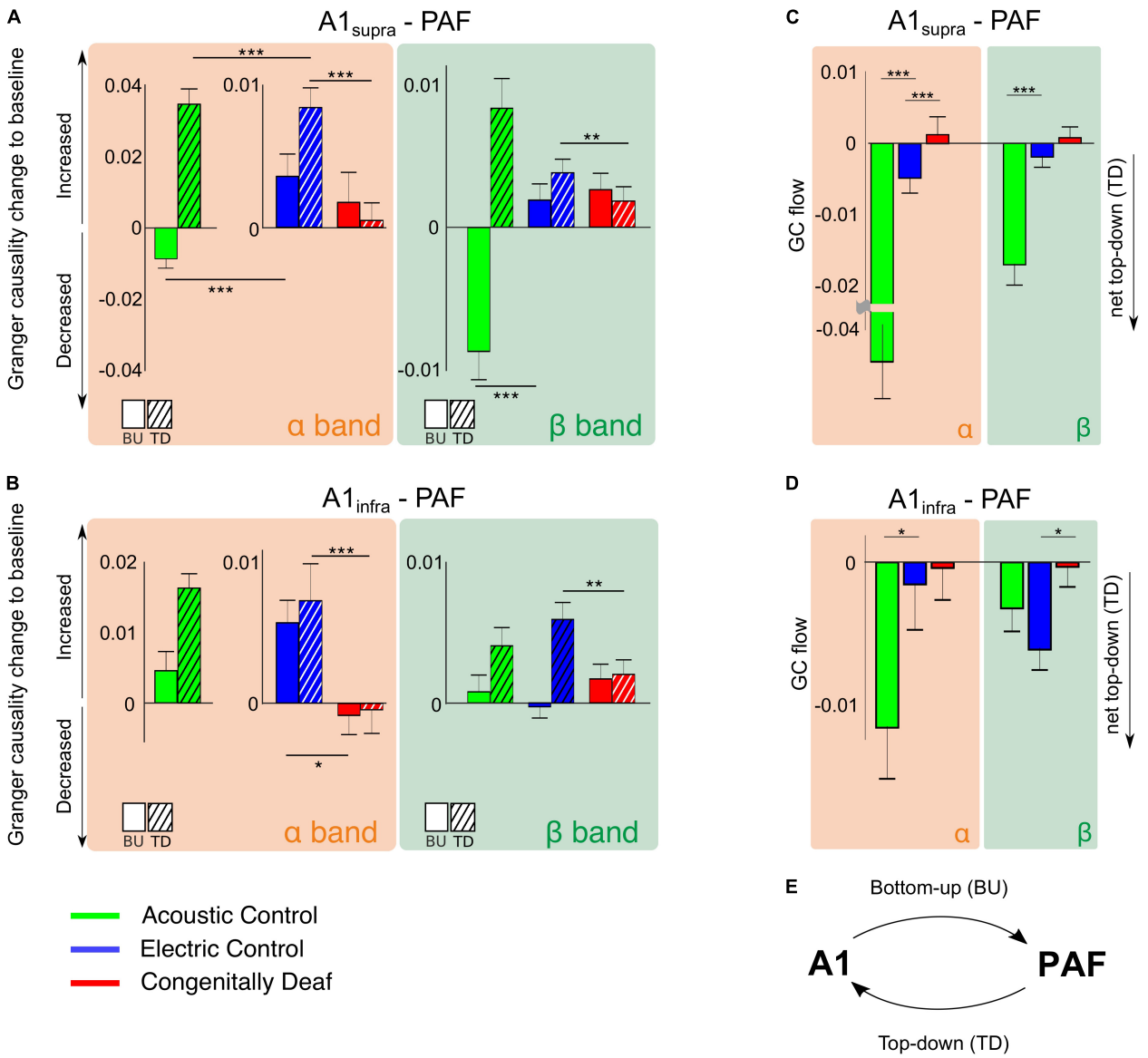

FIGURE 6 | Stimulus-related effective connectivity difference across groups. **(A,B)** Bar plots of alpha and beta band effective connectivity for $A1_{supra}$ – PAF **(A)** and $A1_{infra}$ – PAF **(B)** couplings for all three groups in the late window. Bottom-up (BU) represents connectivity from A1 to PAF while top–down (TD) represents connectivity from PAF to A1. Data shown represent Granger causality (GC) change to baseline (poststimulus GC minus baseline GC). Positive (negative) values indicate increased (decreased) effective connectivity relative to baseline. Solid bars are bottom–up interactions, hatched bars represent top-down interactions. **(C,D)** Barplots of alpha- and beta-band GC flow for $A1_{supra}$ – PAF **(C)** and $A1_{infra}$ – PAF **(D)** couplings for three groups. Data shown are GC flow change to baseline (poststimulus GC flow minus baseline GC flow). GC flow was computed as bottom–up GC minus top–down GC; positive (negative) GC flow values represent domination of bottom–up (top–down) connectivity. Bar plot colors: acoustic control (green), electric control (blue), and congenitally deaf (red) groups. Group pair comparisons were computed using the two-tailed Wilcoxon rank-sum test. $^{*}P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$. **(E)** Illustration of bottom–up (BU) connectivity from A1 to PAF and top–down (TD) connectivity from PAF to A1.

in ACs, had a different coupling pattern, with lower baseline coherence between $A1_{infra}$ and PAF in ACs than in ECs; the effect that was most prominent in the theta band (**Figure 7E**; $p < 0.001$, false discovery rate corrected Wilcoxon rank-sum test).

In congenital deafness, the outcomes differed significantly from ECs particularly for supragranular layers of A1, where CDCs had stronger ongoing coupling to PAF than ECs. This unexpected finding demonstrates that the stimulus-related desynchronization is specific to the auditory stimulus, and further indicates some form of brain adaptation to deafness. Recordings in auditory nerve of CDCs reveal a severely reduced to absent spontaneous activity (Hartmann et al., unpublished observations; for neonatally deafened cats, see Shepherd and Javel, 1997), consequently this rules out a cochlear origin of the difference to ECs.

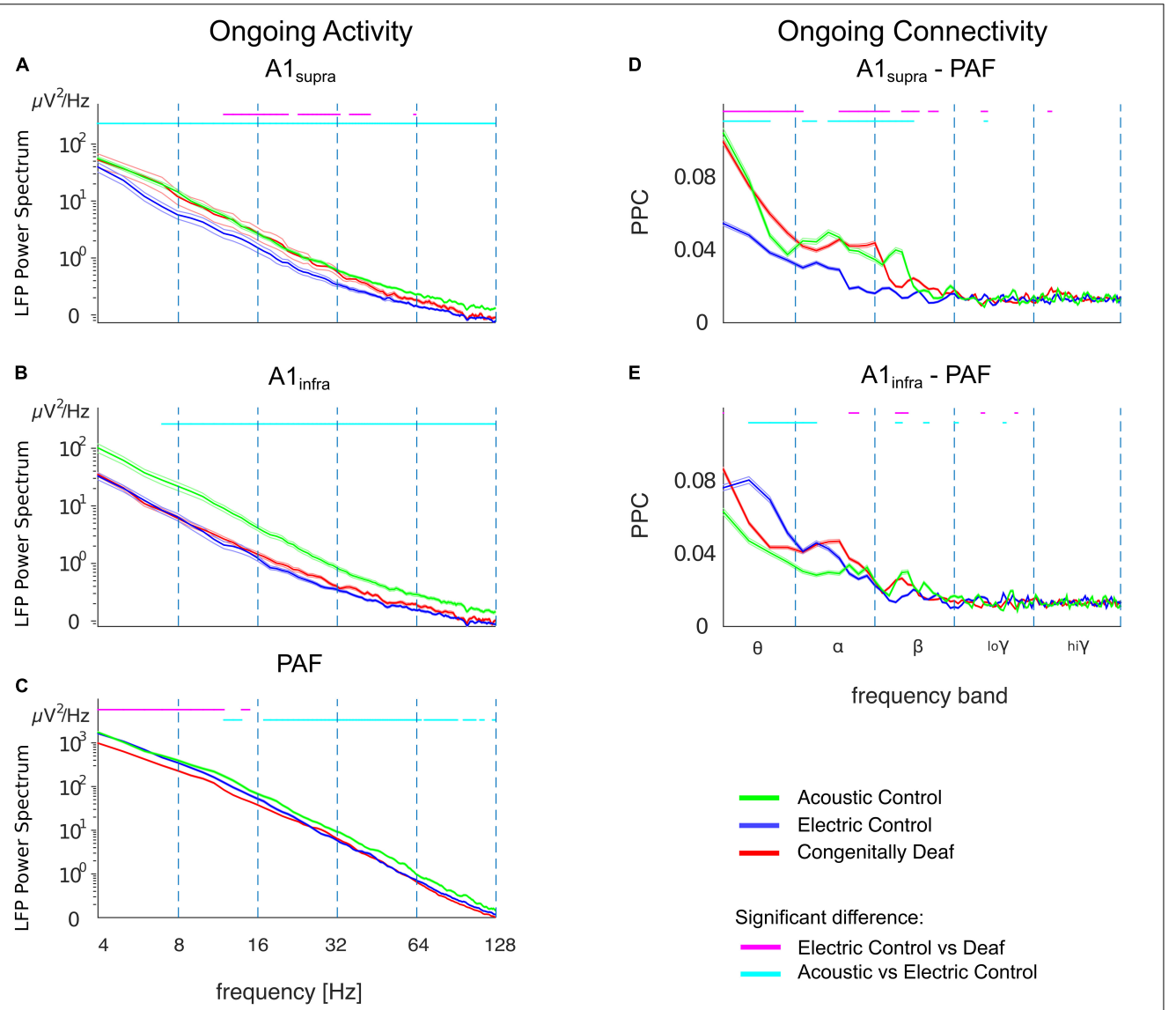

FIGURE 7 | Ongoing LFP power and functional connectivity in A1 and PAF in controls and congenitally deaf cats in absence of stimulation. **(A)** Average LFP power spectra in supragranular A1 ($A1_{supra}$), comparison between acoustic controls with intact cochlea (ACs, green), electric controls with acutely deafened cochlea (ECs, blue), and congenitally deaf cats (CDCs, red) in absence of a stimulus. Data computed from bipolar derivation LFPs. Trials with bursting and spindles were removed before analysis (see section "Materials and Methods"). **(B)** Same as A for infragranular layers in A1 ($A1_{infra}$). **(C)** Average LFP power-spectra in PAF computed from unipolar LFPs. Note the LFP power scale difference. **(D)** $A1_{supra}$ – PAF functional connectivity in absence of stimulation, computed using pairwise phase consistency (PPC). **(E)** Same as D for $A1_{infra}$ – PAF functional connectivity. **(A–E)** Shaded areas represent standard errors of the mean. Statistical pairwise comparisons are shown for electric control vs. deaf (magenta line above the graph) and animals with intact cochleae vs. acutely deafened cochleae (cyan line above the graph) using the two-tailed Wilcoxon rank-sum test (false discovery rate corrected, $p < 0.001$).

It is notable that the ongoing and stimulus-related connectivity revealed different outcomes: whereas in stimulus-related connectivity, CDCs showed weakened couplings between PAF and AI with no sign of synchronization increase in the late time window (**Figures 4E,F**), in ongoing connectivity and supragranular layers they showed a connectivity similar to the ACs (**Figure 7D**). This dissociation demonstrates that the stimulus-related changes are not a mere consequence of ongoing changes. Furthermore it indicates that CDCs partly compensated the effect of absent ongoing drive from the cochlea.

DISCUSSION

The present study directly demonstrates reduced functional and effective stimulus-related connectivity following congenital deafness that is specific to the late processing window (>200 ms post stimulus). Particularly top–down interactions were affected by congenital deafness.

In hearing cats, auditory input synchronized the activity between the areas early in the gamma and beta bands and later (>200 ms) in the alpha and (partly) in the beta band

(**Figure 4**). While in CDCs auditory responses were found in both investigated cortical areas (comp. Yusuf et al., 2017), the stimulus-related coupling between them was significantly weakened in the late window. In contrast to controls, the auditory stimulus predominantly caused interareal desynchronization in CDCs (**Figures 4E,F**). This indicates that the auditory areas of CDCs do not process the stimulus as a functional unit. GC analysis proved that this decoupling mainly reflects reductions in top–down interactions.

The functional and effective connectivity quantifies statistical dependencies between temporal characteristics of neuronal signals (Aertsen and Preissl, 1991; Friston, 2011). Such measures, while faithfully reflecting a functional connection, by definition include the synaptic efficacies of their connections as well as the properties of individual cells and their membranes. All of these affect the ability to form functional connections. The mean evoked LFP response in A1, the consequence of thalamocortical inputs (Lakatos et al., 2009), was not affected by congenital deafness (Kral et al., 2009; Yusuf et al., 2017). This means that the reduced top–down influence from PAF is unlikely due to a downstream effect of a deficient thalamic activation of A1. It might be that bottom-up deficits, either through the weaker A1 → PAF connection, through a weaker thalamic input to PAF, or due to non-reliable responsiveness of PAF neurons to these inputs, could be responsible for the reduced top–down connectivity measures. But in the latter case one would also expect differences in bottom–up connectivity during the early response or in Granger bottom–up results – none of which was the case. The observed reduction of top–down PAF-A1 connectivity was substantially higher than reduction in bottom–up connectivity in the late response in CDCs (**Figure 6**).

Given these considerations, a reduction in bottom–up drive in PAF neurons in CDCs would not be a sufficient explanation of the drop in top–down connectivity. Furthermore, top–down connectivity systematically exceeded the bottom–up connectivity in both hearing groups (**Figure 6**), thus the late processing of the stimulus is normally dominated by top–down influences. This was again not the case in deaf cats. While non-linear effects have to be considered, taken together this suggests that the results faithfully reflect a reduced strength of functional connections between A1 and PAF.

Ongoing activity in CDCs in the supragranular (but not infragranular) layers of A1 coupled, on the other hand, more strongly to PAF. This demonstrates that the auditory areas are not generally decoupled in congenitally deaf; rather, they are specifically decoupled during auditory processing.

Methodology

The approach of accessing fields A1 and PAF and the mapping procedure in A1 in hearing and deaf cats has been validated and described in detail in several previous studies (Kral et al., 2009, 2013; Yusuf et al., 2017). The present results on hearing cats are in line with previous observations of auditory coupling in the auditory cortex of hearing cats, predominantly performed using cross-correlations (Eggermont, 1992, 2000). Previous studies focused on ongoing activity observed a coupling in alpha and beta bands (Eggermont et al., 2011). Auditory correlations increased following an acoustic stimulus (Tomita and Eggermont, 2005), as observed in the present study using phase-based methods (**Figure 4**). To safely prevent bursting from affecting connectivity (Valentine and Eggermont, 2001), we avoided the burst-suppression state and excluded trials with bursts and spindles (see section "Materials and Methods"). The present study, where comparable in hearing controls, is consistent with previous outcomes in hearing cats.

Since the recording sites in A1 and PAF were >1 cm apart (**Figure 1**), volume conduction was unlikely contributing to present results. However, to avoid any volume conduction effects, we used bipolar derivation in field A1. Adopting this approach enabled us also to focus on true synchronization between the recorded sites in absence of signals picked up by the reference electrode. Bipolar derivation also provided local signals and allowed layer-specific analysis in A1. Bipolar derivation was applied only in field A1, since (i) it was sufficient to reliably eliminate the influence of volume conduction and common reference on couplings; (ii) penetrations were perpendicular to cortical layers and the electrical homogeneity of the tissue impedance has been previously shown for this direction (review in Mitzdorf, 1985); (iii) the use of unipolar LFPs in PAF had the advantage of capturing signals from a larger number of PAF neurons, also those localized beyond the track direction, increasing the yield and reducing the dependence on the exact recording location within PAF. The possible drawback is a potential overestimation of the absolute overall connectivity.

We analyzed all data using two phase-based connectivity measures. While debiased weighted phase-lag index (WPLId) is insensitive to volume conduction and thus more sensitive for detecting true connectivity than PPC, the results obtained might be exaggerated depending on the phase angle distribution (review in Cohen, 2014) due to a weighting of the imaginary part of the coherence in the WPLId. PPC, on the other hand, is not biased in phase distributions and is also better comparable to previous outcomes of correlational analyses. Therefore, we used WPLId to identify the significantly coupled site pairs and focused on the PPC in order to analyze their coupling strength (**Figure 4**).

Directionality (effective connectivity) was determined using GC. The results in general corresponded to phase-based measures, but GC additionally showed significant alpha-band difference between AC and EC, which was not observed in PPC. This difference is due to PPC being a symmetric connectivity measure that does not distinguish between bottom–up and top–down influences, considering them aggregately. GC, on the other hand, separated out the information flow and suggested specific decreases and increases in top–down and bottom–up coupling after the stimulus.

To complement the presented coupling analysis with previously used measures of cortical connectivity in hearing cats (Eggermont, 1992, 2000), we additionally performed cross-correlational analysis of the ongoing activity to cross-check our outcomes, obtaining results corresponding to previous studies (results available on request).

Corresponding to previous human data (Hillebrand et al., 2016), also the present study observed a frequency-specificity

in the information flow. The alpha band and to an extent (but less consistently) the beta band played a key role in interareal synchronization following an auditory stimulus at late time windows. The present observations support previous findings highlighting the importance of alpha and beta bands in top-down interactions (Buschman and Miller, 2007; van Kerkoerle et al., 2014; Bastos et al., 2015; Michalareas et al., 2016; Richter et al., 2018). In our study, couplings in the gamma band appeared in the early response (within the first 50 ms post stimulus). The gamma band is considered responsible for bottom–up interactions (Fontolan et al., 2014; Bastos et al., 2015) and did not show any significant differences between the groups. However, the early response is additionally strongly affected by thalamic input to both A1 and PAF (Lee and Winer, 2011). We did not observe strong ongoing synchronization in the gamma band in the late window. Previous studies in humans observed ongoing gamma responses during auditory stimulation (Fontolan et al., 2014). In the present study we used very brief stimuli to avoid the interference from electrical stimulation artifacts, an approach validated in several previous papers (Tillein et al., 2010, 2016; Yusuf et al., 2017). This may have reduced such sustained gamma activity. Human studies, on the other hand, typically used long-duration stimuli that may generate more sustained gamma-oscillations (Ray and Maunsell, 2011) due to the early responses that continue throughout the stimulus. Additionally, gamma transients are often coupled to lower-frequency activity (such as alpha) in the late window (Jensen et al., 2014; Yusuf et al., 2017). When using low-impedance electrodes to record from the cortical surface, these transients may combine from several columns and present as sustained oscillations that we observed only as brief transients with recordings from single columns.

Effects of Anesthesia

Large-scale invasive mapping at dozens of recording positions, including multiple penetrations of the fields, was only possible in anesthetized preparation. The energy of oscillatory phenomena used for coupling quantification is increased by wakefulness, and in particular by attention, but the difference between awake and anesthetized preparations is only quantitative (Fontanini and Katz, 2008; Xing et al., 2012; Sellers et al., 2015), particularly if burst-suppression phenomena are avoided (Land et al., 2012).

Using power-independent measures in the present study eliminated the dependence on signal power, affected by anesthesia. Even when presenting a stimulus passively, it is represented in both primary and secondary fields, and, given this representation, inherently generates both bottom–up and top–down corticocortical interactions, although weaker than in wakefulness and under attention. We obtained significant interareal couplings in both anesthetized control groups. Consequently, while quantitatively stronger coupling can be expected in awake, attentive animals, particularly in top–down interactions (McGinley et al., 2015), and this may yield the statistical comparisons more sensitive, the controls did show significant top–down interactions under anesthesia, and CDCs did not, and the group difference between ECs and CDCs was statistically significant. We can therefore exclude anesthesia as a reason for the differences observed.

Influence of Stimulus Mode (Acoustic vs. Electric)

Granger causality showed weaker top–down connectivity in ECs compared with ACs in all investigated layers of A1 (**Figure 6**). We interpret these observations as a consequence of higher synchrony in CI stimulation and the lack of "naturalness" in the electrical stimulus. The reduced interaction in ECs may thus be the consequence of a stimulus that does not fit into the patterns learned throughout life and stored in auditory cortex (Kral and Eggermont, 2007).

The comparison of acoustic and electric controls allows for differentiation of coupling in condition of a known and unknown stimulus. Predictive coding (Friston, 2010; Keller and Mrsic-Flogel, 2018; Vezoli et al., 2020) assumes that the unknown stimulus not fitting into the patterns stored in higher-order areas would generate a strong bottom-up signal (the prediction error). In supragranular layers in ACs, the strong top–down coupling could be interpreted as a strong prediction and the reduced bottom–up coupling the late window (**Figure 6A**) could be interpreted as a small prediction error. In ECs, the top–down signal (prediction) is smaller and the bottom–up coupling (prediction error) larger. This is consistent with predictive coding. However, in infragranular layers a similar bottom–up signal is observed in both hearing controls. Supragranular layers are the main source of bottom–up stream of information in the cortex, and infragranular layers are the main source of top–down information flow (reviews Hackett, 2011; Markov et al., 2014; Vezoli et al., 2020). That may be the reason why supragranular layers of A1 could better reflect prediction error signaling. However, the present study was not focused on this question and therefore more experiments are required to conclude on this aspect of auditory connectivity.

Layer Specificity

Ongoing functional connectivity between primary supragranular layers and secondary cortex, on the other hand, was upregulated in A1 of deaf animals to the level observed in hearing animals with functional hair cells (**Figure 7D**). This may be related to a general increase in the suprathreshold sensitivity of neurons, as observed in the auditory cortex of congenitally deaf animals (Tillein et al., 2010, 2016). This may partly counterbalance the lack of auditory input.

The higher ongoing functional connectivity between $A1_{supra}$ and PAF in CDCs compared to ECs suggests that the supragranular layers tend to developmentally partially compensate the loss of hair cells. Such increased connectivity could be a reason for the increased baseline LFP power in $A1_{supra}$ of CDCs (**Figure 7A**). In infragranular layers, this phenomenon was not observed, neither in LFP power (**Figure 7B**) nor in functional connectivity (**Figure 7E**).

Layer differences are consistent with previous data reporting reduced activity particularly in deep layers of A1 in CDCs (Kral et al., 2006). Deep layers V and VI have specific function in thalamocorticothalamic loops and thus for auditory stimulus conveyed through the thalamus (de Ribaupierre et al., 1972; Steriade, 1999; Derdikman et al., 2003; Castro-Alamancos,

2004). Layer III, on the other hand, is more related to lateral connections to neighboring columns (Rouiller et al., 1991; Markov et al., 2014). Cytoarchitectonic analysis showed that deep layers (but not supragranular layers) are dystrophic in primary and secondary auditory areas of CDCs (Berger et al., 2017). This is consistent with increased ongoing connectivity observed in supragranular layers of CDCs compared to ECs, since higher supragranular coupling in deaf animals may compensate loss of thalamic input by facilitating lateral propagation of activity within the area A1 (Reimer et al., 2011) but also to PAF – potentially related to cross-modal corticocortical reorganization of this field in CDCs (Lomber et al., 2010). Deep layers, on the other hand, are more closely related to the corresponding thalamic nuclei and thus to processing auditory inputs; these demonstrate more auditory-related deficits.

Development of Corticocortical Connections

Sensory input requires a reciprocal exchange of stimulus-related information at a different level of sensory processing (e.g., different features among each other, or features to object and vice versa) represented in different areas (Malhotra et al., 2004) using interareal couplings (Kral and Sharma, 2012; Kral et al., 2016). This is developmentally shaped by experience that allows for activation, via thalamic inputs, of both primary and secondary auditory areas within a narrow time window of a few milliseconds (**Figure 8**). Firing within such a time window may strengthen the corticocortical synapses that directly connect these areas by processes of synaptic spike-timing dependent plasticity. A developmental process of this nature functionally defines 'auditory' areas by functionally connecting them. Cortical synaptogenesis and synaptic pruning are regulated by hearing experience (Kral et al., 2005). In the absence of auditory input, it may be primarily the synapses that would link different auditory areas during the auditory response that may be excessively pruned (**Figure 8**).

The anatomical connectome was relatively insensitive to developmental modification of experience and cross-modal reorganization (Kral et al., 2003; Lomber et al., 2010; Barone et al., 2013; Land et al., 2016; Butler et al., 2017). A functional shift in auditory areas toward coupling to the visual system in deaf humans has been demonstrated (Bola et al., 2017). Also this effect was observed at the functional level only. The present data demonstrate that it is the auditory functional connectome itself that is extensively shaped by auditory experience.

Consistent with the different involvement of anatomical and functional connectivity in sensory-related effects, our present findings document a dichotomy in the effect of sensory deprivation on ongoing (partly increased, partly decreased) and stimulus-related (decreased) functional connectivity. Thus resting state connectivity, often analyzed in human imaging (magnetic resonance) studies, cannot be equated with stimulus-related connectivity.

Consequences for Cochlear-Implanted Subjects

Sensory inputs are constantly embedded in other brain processing and must "fit" to the processes in higher order areas to propagate there (Kral and Eggermont, 2007). If the acoustic stimulus matches such stimulus templates (priors) stored there, it activates the priors in higher regions, and this results in top–down information flow down to the lower areas that interact with the bottom–up stream (e.g., Keller and Mrsic-Flogel, 2018; Schneider et al., 2018; Vezoli et al., 2020). Top–down interactions play a crucial role in filling-in phenomena and extraction of weak signals in a noisy environment (Davis and Johnsrude, 2007; Petkov et al., 2007; Friston, 2010; Riecke et al., 2012; Wild et al., 2012). Top–down interactions have a crucial role also in entrainment to auditory oscillations (Barczak et al., 2018), for speech understanding (Di Liberto et al., 2018) and also for success of cochlear implantation (Zaltz et al., 2020).

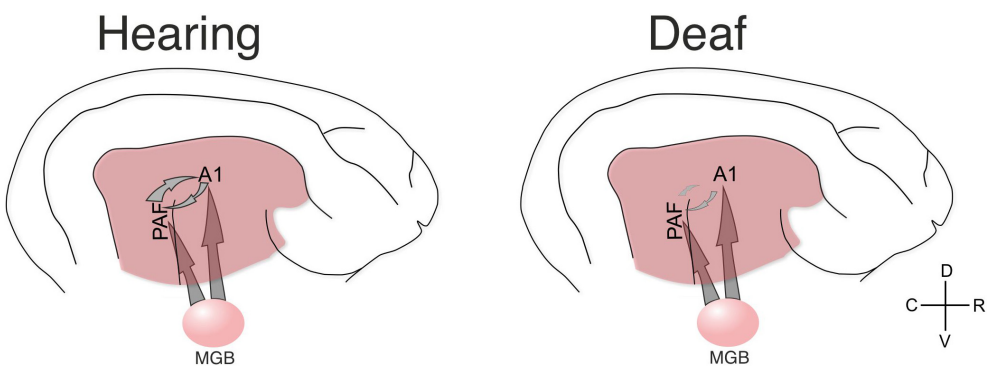

FIGURE 8 | Schematic illustration of the present results. The present data document that in hearing controls the functional connectivity between A1 and PAF is dominated by top–down interactions. Such connectivity is stronger for stimulation that the animal has previous experience with (as shown by comparing acoustic vs. electric stimulation in hearing controls). Both fields additionally receive thalamic input (not directly investigated here) that generates evoked responses and dominates in the early response window. In the deaf animal the functional connectivity in the early window was not significantly different to electric controls, but the late window did not reveal corticocortical synchronization (connectivity) and the dominating influence of top-down information flow disappeared. Yet, responses were observed in both fields, indicating some thalamic input from MGB also in deaf animals (for analysis of the evoked responses, see Yusuf et al., 2017).

The artificial electric stimulus in hearing controls yields significantly weaker top–down connectivity than an acoustic stimulus, most probably since matching with the stimulus priors stored in higher-order areas is poor. At the extreme point, without sensory experience, congenitally deaf animals demonstrate decoupling of connectivity and almost no top–down information flow – likely due to the complete absence of prior internal models. Auditory performance is dependent on bottom–up and top–down interactions during sensory processing (Yusuf et al., 2017), and central processing, executive functioning and "listening strategy" co-determine the benefit of pediatric cochlear implantation (Sharma et al., 2002; Kral et al., 2016). Loss of top–down interactions in the congenitally sensory-deprived brains, normally required for learning control and predictive coding, may be one crucial reason why sensitive periods for therapy of congenital sensory loss eventually close.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The animal study was reviewed and approved by LAVES Oldenburg.

AUTHOR CONTRIBUTIONS

AK designed the project and the experiments and obtained the funding. PH, JT, and AK performed the experiments. PY analyzed the data with supervision of PH, AK, and MV. PY drafted the first version of the manuscript, MV and AK edited it. All the authors approved the manuscript.

FUNDING

This study was supported by the Deutsche Forschungsgemeinschaft (DFG Kr 3370 and Exc 2177), National Science Foundation in cooperation with the German Aerospace Center (DLR 01GQ1703), MedEl Comp, Innsbruck, Austria (to JT), DAAD – Indonesian German Scholarship Programme (IGSP), and Q1Q2 Publication Grant 2019 from Universitas Indonesia (to PY).

ACKNOWLEDGMENTS

We thank Karl-Jürgen Kühne and Daniela Kühne for their helpful assistance in preparation for and during the experiments, and for preparing the histological data. We also thank the reviewers for their comments that improved the manuscript.

REFERENCES

Aertsen, A., and Preissl, H. (1991). "Dynamics of activity and connectivity in physiological neuronal networks," in *Nonlinear Dynamics Neuronal Networks*, ed. H. Schuster (Vancouver: VCH), 281–302.

Avena-Koenigsberger, A., Misic, B., and Sporns, O. (2018). Communication dynamics in complex brain networks. *Nat. Rev. Neurosci.* 19, 17–33. doi: 10.1038/nrn.2017.149

Babapoor-Farrokhran, S., Vinck, M., Womelsdorf, T., and Everling, S. (2017). Theta and beta synchrony coordinate frontal eye fields and anterior cingulate cortex during sensorimotor mapping. *Nat. Commun.* 8:13967. doi: 10.1038/ncomms13967

Barczak, A., O'Connell, M. N., McGinnis, T., Ross, D., Mowery, T., Falchier, A., et al. (2018). Top-down, contextual entrainment of neuronal oscillations in the auditory thalamocortical circuit. *Proc. Natl. Acad. Sci.* 115, E7605–E7614. doi: 10.1073/pnas.1714684115

Barone, P., Lacassagne, L., and Kral, A. (2013). Reorganization of the connectivity of cortical field DZ in congenitally deaf cat. *PLoS One* 8:e60093. doi: 10.1371/journal.pone.0060093

Bastos, A. M., Vezoli, J., Bosman, C. A., Schoffelen, J.-M., Oostenveld, R., Dowdall, J. R., et al. (2015). Visual areas exert feedforward and feedback influences through distinct frequency channels. *Neuron* 85, 390–401. doi: 10.1016/j.neuron.2014.12.018

Bastos, A. M., and Schoffelen, J. (2016). A tutorial review of functional connectivity analysis methods and their interpretational pitfalls. *Front. Syst. Neurosci.* 9, 1–23. doi: 10.3389/fnsys.2015.00175

Benjamini, Y., and Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Ann. Stat.* 29, 1165–1188.

Berger, C., Kühne, D., Scheper, V., and Kral, A. (2017). Congenital deafness affects deep layers in primary and secondary auditory cortex. *J. Comp. Neurol.* 525, 3110–3125. doi: 10.1002/cne.24267

Bizley, J. K., and Cohen, Y. E. (2013). The what, where and how of auditory-object perception. *Nat. Rev. Neurosci.* 14, 693–707. doi: 10.1038/nrn3565

Bola, Ł, Zimmermann, M., Mostowski, P., Jednoróg, K., Marchewka, A., Rutkowski, P., et al. (2017). Task-specific reorganization of the auditory cortex in deaf humans. *Proc. Natl. Acad. Sci.* 114, E600–E609. doi: 10.1073/pnas.1609000114

Busby, P. A., and Clark, G. M. (1999). Gap detection by early-deafened cochlear-implant subjects. *J. Acoust. Soc. Am.* 105, 1841–1852. doi: 10.1121/1.426721

Buschman, T. J., and Miller, E. K. (2007). Top-down versus bottom-up control of attention in the prefrontal and posterior parietal cortices. *Science* 315, 1860–1862. doi: 10.1126/science.1138071

Butler, B. E., Chabot, N., Kral, A., and Lomber, S. G. (2017). Origins of thalamic and cortical projections to the posterior auditory field in congenitally deaf cats. *Hear. Res.* 343, 118–127. doi: 10.1016/j.heares.2016.06.003

Buzsáki, G. (2009). *Rhythms of the Brain*. Oxford: Oxford University Press.

Castro-Alamancos, M. A. (2004). Dynamics of sensory thalamocortical synaptic networks during information processing states. *Prog. Neurobiol.* 74, 213–247. doi: 10.1016/j.pneurobio.2004.09.002

Chang, E. F., and Merzenich, M. M. (2003). Environmental noise retards auditory cortical development. *Science* 300, 498–502. doi: 10.1126/science.1082163

Cohen, M. X. (2014). *Analyzing Neural Time Series Data: Theory and Practice*. Massachusetts: The MIT Press.

Davis, M. H., and Johnsrude, I. S. (2007). Hearing speech sounds: top-down influences on the interface between audition and speech perception. *Hear. Res.* 29, 132–147. doi: 10.1016/j.heares.2007.01.014

de Ribaupierre, F., Goldstein, M. H., and Yeni-Komshian, G. (1972). Intracellular study of the cat's primary auditory cortex. *Brain Res.* 48, 185–204. doi: 10.1016/0006-8993(72)90178-3

Derdikman, D., Hildesheim, R., Ahissar, E., Arieli, A., and Grinvald, A. (2003). Imaging spatiotemporal dynamics of surround inhibition in the barrels somatosensory cortex. *J. Neurosci.* 23, 3100–3105. doi: 10.1523/jneurosci.23-08-03100.2003

Dhamala, M., Rangarajan, G., and Ding, M. (2008). Analyzing information flow in brain networks with nonparametric granger causality. *Neuroimage* 41, 354–362. doi: 10.1016/j.neuroimage.2008.02.020

Di Liberto, G. M., Lalor, E. C., and Millman, R. E. (2018). Causal cortical dynamics of a predictive enhancement of speech intelligibility. *Neuroimage* 166, 247–258. doi: 10.1016/j.neuroimage.2017.10.066

Donner, T. H., and Siegel, M. (2011). A framework for local cortical oscillation patterns. *Trends Cogn. Sci.* 15, 191–199. doi: 10.1016/j.tics.2011.03.007

Eggermont, J. J. (1992). Neural interaction in cat primary auditory cortex. dependence on recording depth, electrode separation, and age. *J. Neurophysiol.* 68, 1216–1228. doi: 10.1152/jn.1992.68.4.1216

Eggermont, J. J. (2000). Sound-induced synchronization of neural activity between and within three auditory cortical areas. *J. Neurophysiol.* 83, 2708–2722. doi: 10.1152/jn.2000.83.5.2708

Eggermont, J. J., Munguia, R., Pienkowski, M., and Shaw, G. (2011). Comparison of LFP-based and spike-based spectro-temporal receptive fields and cross-correlation in cat primary auditory cortex. *PLoS One* 6:e20046. doi: 10.1371/journal.pone.0020046

Fallon, J. B., Irvine, D. R. F., and Shepherd, R. K. (2009). Neural prostheses and brain plasticity. *J. Neural Eng.* 6:065008. doi: 10.1088/1741-2560/6/6/065008

Fontanini, A., and Katz, D. B. (2008). Behavioral states, network states, and sensory response variability. *J. Neurophysiol.* 100, 1160–1168. doi: 10.1152/jn.90592.2008

Fontolan, L., Morillon, B., Liegeois-Chauvel, C., and Giraud, A.-L. (2014). The contribution of frequency-specific activity to hierarchical information processing in the human auditory cortex. *Nat. Commun.* 5:4694. doi: 10.1038/ncomms5694

Fries, P. (2005). A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. *Trends Cogn. Sci.* 9, 474–480. doi: 10.1016/j.tics.2005.08.011

Friston, K. J. (2010). The free-energy principle: a unified brain theory? *Nat. Rev. Neurosci.* 11, 127–138. doi: 10.1038/nrn2787

Friston, K. J. (2011). Functional and effective connectivity: a review. *Brain Connect.* 1, 13–36. doi: 10.1089/brain.2011.0008

Geweke, J. (1982). Measurement of linear dependence and feedback between multiple time series. *J. Am. Stat. Assoc.* 77, 304–313. doi: 10.1080/01621459.1982.10477803

Hackett, T. A. (2011). Information flow in the auditory cortical network. *Hear. Res.* 271, 133–146. doi: 10.1016/j.heares.2010.01.011

Heid, S., Hartmann, R., and Klinke, R. (1998). A model for prelingual deafness, the congenitally deaf white cat–population statistics and degenerative changes. *Hear. Res.* 115, 101–112. doi: 10.1016/S0378-5955(97)00182-182

Hillebrand, A., Tewarie, P., van Dellen, E., Yu, M., Carbo, E. W. S., Douw, L., et al. (2016). Direction of information flow in large-scale resting-state networks is frequency-dependent. *Proc. Natl. Acad. Sci.* 113, 3867–3872. doi: 10.1073/pnas.1515657113

Hipp, J. F., Engel, A. K., and Siegel, M. (2011). Oscillatory synchronization in large-scale cortical networks predicts perception. *Neuron* 69, 387–396. doi: 10.1016/j.neuron.2010.12.027

Jensen, O., Gips, B., Bergmann, T. O., and Bonnefond, M. (2014). Temporal coding organized by coupled alpha and gamma oscillations prioritize visual processing. *Trends Neurosci.* 37, 357–369. doi: 10.1016/j.tins.2014.04.001

Keller, G. B., and Mrsic-Flogel, T. D. (2018). Predictive processing: a canonical cortical computation. *Neuron* 100, 424–435. doi: 10.1016/j.neuron.2018.10.003

Klinke, R., Kral, A., Heid, S., Tillein, J., and Hartmann, R. (1999). Recruitment of the auditory cortex in congenitally deaf cats by long-term cochlear electrostimulation. *Science* 285, 1729–1733. doi: 10.1126/science.285.5434.1729

Kornblith, S., Buschman, T. J., and Miller, E. K. (2016). Stimulus load and oscillatory activity in higher cortex. *Cereb. Cortex* 26, 3772–3784. doi: 10.1093/cercor/bhv182

Kral, A., and Eggermont, J. J. (2007). What's to lose and what's to learn: development under auditory deprivation, cochlear implants and limits of cortical plasticity. *Brain Res. Rev.* 56, 259–269. doi: 10.1016/j.brainresrev.2007.07.021

Kral, A., and Lomber, S. G. (2015). Deaf white cats. *Curr. Biol.* 25, R351–R353. doi: 10.1016/j.cub.2015.02.040

Kral, A., and O'Donoghue, G. M. (2010). Profound deafness in childhood. *N. Engl. J. Med.* 363, 1438–1450. doi: 10.1056/NEJMra0911225

Kral, A., and Sharma, A. (2012). Developmental neuroplasticity after cochlear implantation. *Trends Neurosci.* 35, 111–122. doi: 10.1016/j.tins.2011.09.004

Kral, A., Dorman, M. F., and Wilson, B. S. (2019). Neuronal development of hearing and language: cochlear implants and critical periods. *Annu. Rev. Neurosci.* 42, 47–65. doi: 10.1146/annurev-neuro-080317-061513

Kral, A., Hubka, P., Heid, S., and Tillein, J. (2013). Single-sided deafness leads to unilateral aural preference within an early sensitive period. *Brain* 136, 180–193. doi: 10.1093/brain/aws305

Kral, A., Kronenberger, W. G., Pisoni, D. B., and O'Donoghue, G. M. (2016). Neurocognitive factors in sensory restoration of early deafness: a connectome model. *Lancet Neurol.* 15, 610–621. doi: 10.1016/s1474-4422(16)00034-x

Kral, A., Schröder, J.-H., Klinke, R., and Engel, A. K. (2003). Absence of cross-modal reorganization in the primary auditory cortex of congenitally deaf cats. *Exp. Brain Res.* 153, 605–613. doi: 10.1007/s00221-003-1609-z

Kral, A., Tillein, J., Heid, S., Hartmann, R., and Klinke, R. (2005). Postnatal cortical development in congenital auditory deprivation. *Cereb. Cortex* 15, 552–562. doi: 10.1093/cercor/bhh156

Kral, A., Tillein, J., Heid, S., Klinke, R., and Hartmann, R. (2006). Cochlear implants: cortical plasticity in congenital deprivation. *Prog. Brain Res.* 157, 283–313. doi: 10.1016/S0079-6123(06)57018-57019

Kral, A., Tillein, J., Hubka, P., Schiemann, D., Heid, S., Hartmann, R., et al. (2009). Spatiotemporal patterns of cortical activity with bilateral cochlear implants in congenital deafness. *J. Neurosci.* 29, 811–827. doi: 10.1523/jneurosci.2424-08.2009

Lakatos, P., O'Connell, M. N., Barczak, A., Mills, A., Javitt, D. C., and Schroeder, C. E. (2009). The leading sense: supramodal control of neurophysiological context by attention. *Neuron* 64, 419–430. doi: 10.1016/j.neuron.2009.10.014

Land, R., Baumhoff, P., Tillein, J., Lomber, S. G., Hubka, P., and Kral, A. (2016). Cross-modal plasticity in higher-order auditory cortex of congenitally deaf cats does not limit auditory responsiveness to cochlear implants. *J. Neurosci.* 36, 6175–6185. doi: 10.1523/jneurosci.0046-16.2016

Land, R., Engler, G., Kral, A., and Engel, A. K. (2012). Auditory evoked bursts in mouse visual cortex during isoflurane anesthesia. *PLoS One* 7:e49855. doi: 10.1371/journal.pone.0049855

Lee, C. C., and Winer, J. A. (2011). Convergence of thalamic and cortical pathways in cat auditory cortex. *Hear. Res.* 274, 85–94. doi: 10.1016/j.heares.2010.05.008

Lee, C.-C., and Middlebrooks, J. C. (2013). Specialization for sound localization in fields A1, DZ, and PAF of cat auditory cortex. *J. Assoc. Res. Otolaryngol.* 14, 61–82. doi: 10.1007/s10162-012-0357-9

Lomber, S. G., and Malhotra, S. (2008). Double dissociation of "what" and "where" processing in auditory cortex. *Nat. Neurosci.* 11, 609–616. doi: 10.1038/nn.2108

Lomber, S. G., Meredith, M. A., and Kral, A. (2010). Cross-modal plasticity in specific auditory cortices underlies visual compensations in the deaf. *Nat. Neurosci.* 13, 1421–1427. doi: 10.1038/nn.2653

Malhotra, S., Hall, A. J., and Lomber, S. G. (2004). Cortical control of sound localization in the cat: unilateral cooling deactivation of 19 cerebral areas. *J. Neurophysiol.* 92, 1625–1643. doi: 10.1152/jn.01205.2003

Manrique, M., Cervera-Paz, F. J., Huarte, A., Perez, N., Molina, M., and Garcıìa-Tapia, R. (1999). Cerebral auditory plasticity and cochlear implants. *Int. J. Pediatr. Otorhinolaryngol.* 49, S193–S197.

Maris, E., and Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *J. Neurosci. Methods* 164, 177–190. doi: 10.1016/j.jneumeth.2007.03.024

Markov, N. T., Vezoli, J., Chameau, P., Falchier, A., Quilodran, R., Huissoud, C., et al. (2014). Anatomy of hierarchy: feedforward and feedback pathways in macaque visual cortex. *J. Comp. Neurol.* 522, 225–259. doi: 10.1002/cne.23458

McGinley, M. J., Vinck, M., Reimer, J., Batista-Brito, R., Zagha, E., Cadwell, C. R., et al. (2015). Waking state: rapid variations modulate neural and behavioral responses. *Neuron* 87, 1143–1161. doi: 10.1016/j.neuron.2015.09.012

Michalareas, G., Vezoli, J., van Pelt, S., Schoffelen, J.-M., Kennedy, H., and Fries, P. (2016). Alpha-beta and gamma rhythms subserve feedback and feedforward influences among human visual cortical areas. *Neuron* 89, 384–397. doi: 10.1016/j.neuron.2015.12.018

Mitani, A., and Shimokouchi, M. (1985). Neuronal connections in the primary auditory cortex: an electrophysiological study in the cat. *J. Comp. Neurol.* 235, 417–429. doi: 10.1002/cne.902350402

Mitzdorf, U. (1985). Current source-density method and application in cat cerebral cortex: investigation of evoked potentials and EEG phenomena. *Physiol. Rev.* 65, 37–100. doi: 10.1152/physrev.1985.65.1.37

Niparko, J. K., Tobey, E. A., Thal, D. J., Eisenberg, L. S., Wang, N.-Y., Quittner, A. L., et al. (2010). Spoken language development in children following cochlear implantation. *JAMA* 303, 1498–1506. doi: 10.1001/jama.2010.451

Nolte, G., Ziehe, A., Nikulin, V. V., Schlögl, A., Krämer, N., Brismar, T., et al. (2008). Robustly estimating the flow direction of information in complex physical systems. *Phys. Rev. Lett.* 100:234101.

Oostenveld, R., Fries, P., Maris, E., and Schoffelen, J.-M. (2011). FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput. Intell. Neurosci.* 2011:156869. doi: 10.1155/2011/156869

Petkov, C. I., O'Connor, K. N., and Sutter, M. L. (2007). Encoding of illusory continuity in primary auditory cortex. *Neuron* 54, 153–165. doi: 10.1016/j.neuron.2007.02.031

Phillips, J. M., Vinck, M., Everling, S., and Womelsdorf, T. (2014). A long-range fronto-parietal 5- to 10-Hz network predicts "top-down" controlled guidance in a task-switch paradigm. *Cereb. Cortex* 24, 1996–2008. doi: 10.1093/cercor/bht050

Ponton, C. W., and Eggermont, J. J. (2001). Of kittens and kids: altered cortical maturation following profound deafness and cochlear implant use. *Audiol. Neurootology.* 6, 363–380. doi: 10.1159/000046846

Rauschecker, J. P. (1995). Compensatory plasticity and sensory substitution in the cerebral cortex. *Trends Neurosci.* 18, 36–43. doi: 10.1016/0166-2236(95)93948-w

Rauschecker, J. P., and Tian, B. (2000). Mechanisms and streams for processing of "what" and "where" in auditory cortex. *Proc. Natl. Acad. Sci.* 97, 11800–11806. doi: 10.1073/pnas.97.22.11800

Ray, S., and Maunsell, J. H. R. (2011). Different origins of gamma rhythm and high-gamma activity in macaque visual cortex. *PLoS Biol.* 9:e1000610. doi: 10.1371/journal.pbio.1000610

Reimer, A., Hubka, P., Engel, A. K., and Kral, A. (2011). Fast propagating waves within the rodent auditory cortex. *Cereb. Cortex* 21, 166–177. doi: 10.1093/cercor/bhq073

Richter, C. G., Coppola, R., and Bressler, S. L. (2018). Top-down beta oscillatory signaling conveys behavioral context in early visual cortex. *Sci. Rep.* 8:6991.

Richter, C. G., Thompson, W. H., Bosman, C. A., and Fries, P. (2015). A jackknife approach to quantifying single-trial correlation between covariance-based metrics undefined on a single-trial basis. *Neuroimage* 114, 57–70. doi: 10.1016/j.neuroimage.2015.04.040

Riecke, L., Vanbussel, M., Hausfeld, L., Baskent, D., Formisano, E., and Esposito, F. (2012). Hearing an illusory vowel in noise: suppression of auditory cortical activity. *J. Neurosci.* 31, 8024–8034. doi: 10.1523/jneurosci.0440-12.2012

Rouiller, E. M., Simm, G. M., Villa, A. E. P., de Ribaupierre, Y., and de Ribaupierre, F. (1991). Auditory corticocortical interconnections in the cat: evidence for parallel and hierarchical arrangement of the auditory cortical areas. *Exp. Brain Res.* 86, 483–505. doi: 10.1007/BF00230523

Sato, M., Baumhoff, P., and Kral, A. (2016). Cochlear implant stimulation of a hearing ear generates separate electrophonic and electroneural responses. *J. Neurosci.* 36, 54–64. doi: 10.1523/jneurosci.2968-15.2016

Sato, M., Baumhoff, P., Tillein, J., and Kral, A. (2017). Physiological mechanisms in combined electric–acoustic stimulation. *Otol. Neurotol.* 38, e215–e223.

Schneider, D. M., Sundararajan, J., and Mooney, R. (2018). A cortical filter that learns to suppress the acoustic consequences of movement. *Nature* 561, 391–395. doi: 10.1038/s41586-018-0520-5

Sellers, K. K., Bennett, D. V., Hutt, A., Williams, J. H., and Fröhlich, F. (2015). Awake vs. anesthetized: layer-specific sensory processing in visual cortex and functional connectivity between cortical areas. *J. Neurophysiol.* 113, 3798–3815. doi: 10.1152/jn.00923.2014

Sharma, A., Dorman, M. F., and Kral, A. (2005). The influence of a sensitive period on central auditory development in children with unilateral and bilateral cochlear implants. *Hear. Res.* 203, 134–143. doi: 10.1016/j.heares.2004.12.010

Sharma, A., Dorman, M. F., and Spahr, A. J. (2002). A sensitive period for the development of the central auditory system in children with cochlear implants: implications for age of implantation. *Ear Hear.* 23, 532–539. doi: 10.1097/00003446-200212000-00004

Shepherd, R. K., and Javel, E. (1997). Electrical stimulation of the auditory nerve. i. correlation of physiological responses with cochlear status. *Hear. Res.* 108, 112–144. doi: 10.1016/s0378-5955(97)00046-4

Siegel, M., Donner, T. H., and Engel, A. K. (2012). Spectral fingerprints of large-scale neuronal interactions. *Nat. Rev. Neurosci.* 13, 20–25. doi: 10.1038/nrn3137

Stecker, G. C., Mickey, B. J., Macpherson, E. A., and Middlebrooks, J. C. (2003). Spatial sensitivity in field PAF of cat auditory cortex. *J. Neurophysiol.* 89, 2889–2903. doi: 10.1152/jn.00980.2002

Steriade, M. (1999). Coherent oscillations and short-term plasticity in corticothalamic networks. *Trends Neurosci.* 22, 337–345. doi: 10.1016/s0166-2236(99)01407-1

Suárez, L. E., Markello, R. D., Betzel, R. F., and Misic, B. (2020). Linking structure and function in macroscale brain networks. *Trends Cogn. Sci.* 24, 302–315. doi: 10.1016/j.tics.2020.01.008

Tillein, J., Hubka, P., and Kral, A. (2016). Monaural congenital deafness affects aural dominance and degrades binaural processing. *Cereb. Cortex* 26, 1762–1777. doi: 10.1093/cercor/bhv351

Tillein, J., Hubka, P., Syed, E., Hartmann, R., Engel, A. K., and Kral, A. (2010). Cortical representation of interaural time difference in congenital deafness. *Cereb. Cortex* 20, 492–506. doi: 10.1093/cercor/bhp222

Tomita, M., and Eggermont, J. J. (2005). Cross-correlation and joint spectro-temporal receptive field properties in auditory cortex. *J. Neurophysiol.* 93, 378–392. doi: 10.1152/jn.00643.2004

Valentine, P. A., and Eggermont, J. J. (2001). Spontaneous burst-firing in three auditory cortical fields: its relation to local field potentials and its effect on inter-area cross-correlations. *Hear. Res.* 154, 146–157. doi: 10.1016/s0378-5955(01)00241-6

van Kerkoerle, T., Self, M. W., Dagnino, B., Gariel-Mathis, M.-A., Poort, J., van der Togt, C., et al. (2014). Alpha and gamma oscillations characterize feedback and feedforward processing in monkey visual cortex. *Proc. Natl. Acad. Sci.* 111, 14332–14341. doi: 10.1073/pnas.1402773111

Vezoli, J., Magrou, L., Goebel, R., Wang, X. J., Knoblauch, K., Vinck, M., et al. (2020). Cortical hierarchy, dual counterstream architecture and the importance of top-down generative networks. *Neuroimage* 225:117479. doi: 10.1016/j.neuroimage.2020.117479

Vinck, M., Huurdeman, L., Bosman, C. A., Fries, P., Battaglia, F. P., Pennartz, C. M. A., et al. (2015). How to detect the Granger-causal flow direction in the presence of additive noise? *Neuroimage* 108, 301–318. doi: 10.1016/j.neuroimage.2014.12.017

Vinck, M., Oostenveld, R., Van Wingerden, M., Battaglia, F. P., and Pennartz, C. M. A. (2011). An improved index of phase-synchronization for electrophysiological data in the presence of volume-conduction, noise and sample-size bias. *Neuroimage* 55, 1548–1565. doi: 10.1016/j.neuroimage.2011.01.055

Vinck, M., van Wingerden, M., Womelsdorf, T., Fries, P., and Pennartz, C. M. A. (2010). The pairwise phase consistency: a bias-free measure of rhythmic neuronal synchronization. *Neuroimage* 51, 112–122. doi: 10.1016/j.neuroimage.2010.01.073

Wei, C., Cao, K., Jin, X., Chen, X., and Zeng, F.-G. (2007). Psychophysical performance and mandarin tone recognition in noise by cochlear implant users. *Ear Hear.* 28(2 Suppl.), 62S–65S.

Wild, C. J., Yusuf, A., Wilson, D. E., Peelle, J. E., Davis, M. H., and Johnsrude, I. S. (2012). Effortful listening: the processing of degraded speech depends critically on attention. *J. Neurosci.* 32, 14010–14021. doi: 10.1523/jneurosci.1528-12.2012

Winer, J. A., and Lee, C. C. (2007). The distributed auditory cortex. *Hear. Res.* 229, 3–13.

Womelsdorf, T., Schoffelen, J.-M., Oostenveld, R., Singer, W., Desimone, R., Engel, A. K., et al. (2007). Modulation of neuronal interactions through neuronal synchronization. *Science* 316, 1609–1612. doi: 10.1126/science.1139597

Xing, D., Shen, Y., Burns, S., Yeh, C.-I., Shapley, R. M., and Li, W. (2012). Stochastic generation of gamma-band activity in primary visual cortex of awake and anesthetized monkeys. *J. Neurosci.* 32:13873–80a.

Yusuf, P. A., Hubka, P., Tillein, J., and Kral, A. (2017). Induced cortical responses require developmental sensory experience. *Brain* 140, 3153–3165. doi: 10.1093/brain/awx286

Zaltz, Y., Bugannim, Y., Zechoval, D., Kishon-Rabin, L., and Perez, R. (2020). Listening in noise remains a significant challenge for cochlear implant users: evidence from early deafened and those with progressive hearing loss compared to peers with normal hearing. *J. Clin. Med.* 9:1381. doi: 10.3390/jcm9051381

Conflict of Interest: JT was employed by MedEl Company, Innsbruck, Austria.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Yusuf, Hubka, Tillein, Vinck and Kral. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 16 February 2021
doi: 10.3389/fnins.2021.640127

Correlation and Reliability of Behavioral and Otoacoustic-Emission Estimates of Contralateral Medial Olivocochlear Reflex Strength in Humans

*Miriam I. Marrufo-Pérez[1,2], Peter T. Johannesen[1,2] and Enrique A. Lopez-Poveda[1,2,3]**

[1] Instituto de Neurociencias de Castilla y León, Universidad de Salamanca, Salamanca, Spain, [2] Instituto de Investigación Biomédica de Salamanca, Universidad de Salamanca, Salamanca, Spain, [3] Departamento de Cirugía, Facultad de Medicina, Universidad de Salamanca, Salamanca, Spain

The roles of the medial olivocochlear reflex (MOCR) in human hearing have been widely investigated but remain controversial. We reason that this may be because the effects of MOCR activation on cochlear mechanical responses can be assessed only indirectly in healthy humans, and the different methods used to assess those effects possibly yield different and/or unreliable estimates. One aim of this study was to investigate the correlation between three methods often employed to assess the strength of MOCR activation by contralateral acoustic stimulation (CAS). We measured tone detection thresholds (N = 28), click-evoked otoacoustic emission (CEOAE) input/output (I/O) curves (N = 18), and distortion-product otoacoustic emission (DPOAE) I/O curves (N = 18) for various test frequencies in the presence and the absence of CAS (broadband noise of 60 dB SPL). As expected, CAS worsened tone detection thresholds, suppressed CEOAEs and DPOAEs, and horizontally shifted CEOAE and DPOAE I/O curves to higher levels. However, the CAS effect on tone detection thresholds was not correlated with the horizontal shift of CEOAE or DPOAE I/O curves, and the CAS-induced CEOAE suppression was not correlated with DPOAE suppression. Only the horizontal shifts of CEOAE and DPOAE I/O functions were correlated with each other at 1.5, 2, and 3 kHz. A second aim was to investigate which of the methods is more reliable. The test–retest variability of the CAS effect was high overall but smallest for tone detection thresholds and CEOAEs, suggesting that their use should be prioritized over the use of DPOAEs. Many factors not related with the MOCR, including the limited parametric space studied, the low resolution of the I/O curves, and the reduced numbers of observations due to data exclusion likely contributed to the weak correlations and the large test–retest variability noted. These findings can help us understand the inconsistencies among past studies and improve our understanding of the functional significance of the MOCR.

Keywords: basilar membrane, suppression, olivocochlear efferents, effective attenuation, input/output curves, contralateral acoustic stimulation

OPEN ACCESS

Edited by:
David K. Ryugo, Garvan Institute of Medical Research, Australia

Reviewed by:
Paul Hinckley Delano, University of Chile, Chile
Carolina Abdala, University of Southern California, United States

****Correspondence:***
Enrique A. Lopez-Poveda
ealopezpoveda@usal.es

Specialty section:
This article was submitted to Auditory Cognitive Neuroscience, a section of the journal Frontiers in Neuroscience

Received: *10 December 2020*
Accepted: *26 January 2021*
Published: *16 February 2021*

Citation:
Marrufo-Pérez MI, Johannesen PT and Lopez-Poveda EA (2021) Correlation and Reliability of Behavioral and Otoacoustic-Emission Estimates of Contralateral Medial Olivocochlear Reflex Strength in Humans. Front. Neurosci. 15:640127. doi: 10.3389/fnins.2021.640127

INTRODUCTION

The central nervous system can adjust the functioning of the inner ear via the olivocochlear efferent system. Some efferent fibers originate in the medial region of the superior olivary complex and terminate on the outer hair cells (OHCs) in the cochlea (Warr and Guinan, 1979). These fibers, termed medial olivocochlear (MOC) efferents, can be activated reflexively by sounds presented to the ipsilateral and/or the contralateral ear (Liberman and Brown, 1986; Brown et al., 2003). It has been suggested that the MOC reflex (MOCR) serves to protect the auditory system from acoustic overstimulation and to facilitate auditory perception in noise, among other. However, the evidence in support of these roles is mixed (reviewed by Fuente, 2015; Smith and Keil, 2015; Lopez-Poveda, 2018). Because the effects of the MOCR can be assessed only indirectly in healthy humans, the existing evidence is mostly based on correlations between a psychoacoustic measure of interest (e.g., noise-induced temporary threshold shifts or speech-in-noise recognition) and indirect estimates of the inhibition of basilar membrane (BM) responses by MOCR activation, often referred to as MOCR strength. Different studies have used different techniques to estimate MOCR strength. If the different methods yielded uncorrelated or unreliable estimates of MOCR strength, this could partly explain the discrepant findings regarding the roles of the MOCR in human hearing. The aim of the present study was to investigate the correlation and reliability of three different methods often employed to estimate MOCR strength in humans.

Activation of MOC efferents hyperpolarizes OHCs (Cooper and Guinan, 2003), turning down the gain of the cochlear amplifier at low-to-mid levels and linearizing BM input/output (I/O) curves (Murugasu and Russell, 1996; Dolan et al., 1997; Cooper and Guinan, 2003, 2006; Guinan, 2006). For a tone in noise, MOC efferents inhibit the cochlear mechanical response to the noise and tone stimuli. As a result, auditory nerve fibers respond less to the background noise and show less 'compressed' rate-level functions (Winslow and Sachs, 1988; Kawase et al., 1993). Animal studies suggest that MOC efferents can protect the auditory system from acoustic trauma (Handrock and Zeisberg, 1982; Kujawa and Liberman, 1997; Maison and Liberman, 2000; Maison et al., 2013) and/or enhance the neural representation of transient stimuli in noisy backgrounds (Nieder and Nieder, 1970a,b). However, the results from human studies are not always consistent with these notions (Fuente, 2015; Lopez-Poveda, 2018).

In animals, the roles of MOC efferents have been studied by interrupting or sectioning the MOCR pathways (e.g., Handrock and Zeisberg, 1982; Warren and Liberman, 1989; Kujawa and Liberman, 1997; Maison et al., 2013). This approach is not always feasible in humans and vestibular neurectomy (the procedure employed to section olivocochlear efferents) is likely ineffective in cutting all olivocochlear efferents (Chays et al., 2003). For these reasons, many human studies have sought to establish a correlation between auditory perceptual tasks hypothesized to depend on the MOCR and an effect of MOCR activation on BM responses. Different methods have been used to assess MOCR effects. For example, many studies have estimated MOCR strength as the level change in click-evoked (CEOAEs) or distortion-product otoacoustic emissions (DPOAEs) induced by contralateral acoustic stimulation (CAS) (e.g., Giraud et al., 1997; Kumar and Vanaja, 2004; De Boer et al., 2012; Stuart and Butler, 2012; Abdala et al., 2014; Mishra and Lutman, 2014; Bidelman and Bhagat, 2015; Mertes et al., 2018, 2019). Because a contralateral broadband noise (BBN) with sufficient level [$\geq$30 dB sound pressure level (SPL); Moulin et al., 1993] activates the contralateral MOCR, and because otoacoustic emissions (OAEs) require OHC-mediated amplification (Shera and Abdala, 2012), the suppression of CEOAEs or DPOAEs by CAS is thought to be the result of the MOCR reducing cochlear gain. MOCR strength has been also estimated as the CAS-induced change in OAE I/O curves (Moulin et al., 1993; Veuillet et al., 1996; Abdala et al., 1999), in behaviorally inferred BM I/O curves (Yasin et al., 2014; Fletcher et al., 2016), and in tone detection thresholds (Kawase et al., 2003; Aguilar et al., 2015; Nogueira et al., 2019).

It is yet to be shown, however, that the different methods used to assess MOCR strength in humans yield reliable and correlated results. In fact, studies aimed at investigating the facilitating role of the MOCR in speech-in-noise recognition have shown discrepant findings when using different methods to assess MOCR strength. For instance, monaural speech reception thresholds (SRTs) for sentences in noise are correlated with CAS-induced CEOAE suppression (Bidelman and Bhagat, 2015) but not with DPOAE suppression (Mukari and Mamat, 2008). Strikingly, findings can be discrepant even when MOCR strength is assessed using the same method. For example, Bidelman and Bhagat (2015) found SRTs for sentences in noise to be correlated with CEOAE suppression, while Stuart and Butler (2012) did not, something remarkable considering that the two studies measured CEOAE suppression using identical stimuli [60 dB peak-equivalent SPL (pSPL) linear clicks at a rate of 50/s and contralateral BBN of 65 SPL]. Mertes et al. (2018) observed a correlation between CAS-induced CEOAE suppression and the slope of the psychometric function for words in noise, but Mertes et al. (2019) did not find such a correlation for the same speech material. Notably, Mertes et al. (2018) measured CEOAEs using 75 dB pSPL clicks while Mertes et al. (2019) used 65 dB pSPL clicks. It is possible that differences across studies in the speech tests or participants contribute to the discrepant findings, but it is also possible that the effects of CAS on CEOAEs and DPOAEs are not reliable or equivalent to assess MOCR strength.

Here, we investigate the correlation and reliability of three popular ways of assessing the strength of the contralateral MOCR in humans. We measured pure-tone detection thresholds at different frequencies as well as CEOAEs and DPOAEs for different test frequencies and levels (i.e., I/O curves). All measures were obtained with and without CAS to compare the "CAS effect" across measures. They were also obtained multiple times to assess the variability of the CAS effect for each measure. Low test–retest variability together with a high correlation of the CAS effect between the different measures would support that the three measures are reliable and consistent, and thus serve equally to

assess MOCR strength. By contrast, high test–retest variability and/or a lack of correlation between methods would indicate that different factors are probably involved in the CAS effects for each measure, which would help to understand the inconsistencies among studies and improve our understanding of the functional significance of the MOCR.

MATERIALS AND METHODS

Participants

Twenty-eight subjects (21 women) with no self-reported history of hearing impairment participated in the study, although not all of them participated in every test (see below). Their mean age was 27.5 years (standard deviation, $SD = 7.5$ years; age range = 18–47 years). Air conduction audiometric thresholds were measured using a clinical audiometer (Interacoustics AD229e). All but three of the participants had air conduction audiometric thresholds $\leq$ 20 dB hearing level (HL) in both ears at frequencies between 125 Hz and 8 kHz (ANSI, 1996). The exceptions were two participants whose threshold was 25 dB HL at 8 kHz in the left and/or right ear, and another participant whose threshold was 60 dB HL at 8 kHz in the right ear. This latter participant was nevertheless admitted for testing because her thresholds were normal over the frequency range of interest for the present study ($\leq$4 kHz). Twenty-six subjects had normal tympanograms (assessed using an Interacoustics AT235h clinical tympanometer and a test tone of 226 Hz at 85 dB SPL). Two listeners had slightly higher than typical values for ear-canal volume, compliance values, and/or tympanic peak pressure in one ear.

Participants were volunteers and not paid for their services.

Tone Detection Thresholds

Absolute detection thresholds in the presence and in the absence of CAS were measured for tones presented monaurally in the left ear of 15 participants and in the right ear of 13 participants ($N = 28$ participants in total). Pure tone frequencies were 0.5, 1.5, and 4 kHz. The duration of the tones was 300 ms, including 10-ms raised-cosine onset and offset ramps. The CAS was a BBN (0.01–10 kHz). This noise bandwidth was used because it produces the greatest MOCR activation (Maison et al., 2000; Lilaonitkul and Guinan, 2009). The CAS level was 60 dB SPL. This level is capable of activating the MOCR with minimal or no activation of the middle-ear muscle reflex (Zhao and Dhar, 2010; Aguilar et al., 2013; Mishra and Lutman, 2013; Mertes and Leek, 2016; Feeney et al., 2017). The CAS had a duration of 850 ms, including 5-ms raised-cosine onset and offset ramps.

A three-interval, three-alternative, forced-choice adaptive procedure was used to measure tone detection thresholds. Three intervals were presented to the listener accompanied by brief lights in a computer monitor, and the tone was presented in one of the intervals chosen at random. The lights were on for 850 ms, and the inter-interval time (the period between the offset and the onset of the lights) was 500 ms. In the conditions with CAS, the CAS was presented in the three intervals gated with the lights. The tone started 500 ms after the light onset in the conditions with and without CAS. That is, the tone started 500 ms after the noise onset in the conditions with CAS. Because the MOCR is almost fully activated about 280 ms after the elicitor onset (Backus and Guinan, 2006), we assumed that the CAS-activated MOCR was fully active at the onset of the tone and remained active over the tone duration.

Participants were instructed to identify the interval containing the tone by pressing a key on the computer keyboard, and feedback was given on the correctness of their responses. The level of the tone decreased after two successive correct responses and increased after an incorrect response (two-down, one-up adaptive rule). The tone detection threshold was thus defined as the tone level giving 70.7% correct responses in the psychometric function (Levitt, 1971). The level of the tone changed by 6 dB until the second reversal in level occurred, and by 2 dB thereafter. The procedure continued until 12 level reversals occurred, and the detection threshold was defined as the mean of the tone levels at the last 10 reversals.

Tone thresholds with and without CAS were always measured in pairs without removing the earphones to avoid measurement variance from the earphones fit, and the threshold without CAS was always measured first. A given pair of thresholds was discarded when the within-measure SD for one or the two thresholds in the pair exceeded 4 dB. The exceptions were three participants for whom we accepted SD $\leq$ 6 dB at 0.5 kHz. Three threshold pairs (with and without CAS) were obtained for each tone frequency. When the across-measure SD of the three thresholds with or without CAS exceeded 4 dB, an additional pair of thresholds was measured. The three (or four) thresholds were averaged and the mean was taken as the tone detection threshold. Thresholds for the three test frequencies were measured in random order across participants.

Click-Evoked Otoacoustic Emissions (CEOAEs)

Click-evoked otoacoustic emissions for the same ear as tone detection thresholds were measured in the presence and in the absence of CAS. CEAOEs were measured using the linear method, in which the responses to four clicks of the same amplitude and polarity were averaged (Kemp et al., 1990). This method was used because although the non-linear method is less sensitive to artifacts, it also cancels linear components of the OAEs and can eliminate much OAEs from the recording (Shera and Abdala, 2012), including the linear part of the MOC effect (Guinan, 2006). For each CEOAE measurement, 1,024 clicks of 75 μs in duration were presented at a rate of 19 Hz. The use of click rates $\leq$25 Hz minimizes the probability of clicks activating the ipsilateral MOCR (Boothalingam and Purcell, 2015). A 19.5 ms response window was used to extract the CEOAE level from the average waveform. The window started 2.5 ms after the end of the click to minimize stimulus artifact. In addition to the overall CEOAE level, the spectrum of the recording was calculated to obtain CEOAE levels at five frequency bands centered at 1, 1.5, 2, 3, and 4 kHz.

Click-evoked otoacoustic emissions for click levels of 51, 54, 57, and 60 dB pSPL[1] were measured in 18, 28, 18, and 20 participants, respectively. In other words, full CEOAE I/O functions (i.e., CEOAEs for the four click levels) were obtained in 18 participants. Eight CEAOE measures (of 1,024 clicks each) were obtained for each click level. Four measures were obtained without CAS and four measures were obtained with CAS. Measurements with and without CAS were interleaved. For any given click level and frequency band, the mean CEOAE level with or without CAS was calculated when at least three of the four pair of measures (with and without CAS) were valid and when the across-measures SD was ≤3 dB both with and without CAS. A measure was regarded as valid when the signal-to-noise ratio (SNR) was ≥6 dB. The mean CEOAE level must be at least 3 dB higher than the system's artifact level to be included in the analyses. If a measure did not meet these criteria, it was classified as "no response."

The CAS had the same characteristics as described for tone detection thresholds, with the exception of its duration. Here, the CAS onset and offset were controlled manually by the experimenter. The CAS started one-to-two seconds before the presentation of the first click and was continuously on until one-to-two seconds after the presentation of the last click.

Distortion-Product Otoacoustic Emissions (DPOAEs)

For 18 participants, $2f_1 - f_2$ DPOAEs were measured in the same ear as tone detection thresholds in the presence and the absence of CAS. The primary f_2 frequencies were 1, 1.5, 2, 3, and 4 kHz, and the f_2/f_1 ratio was fixed at 1.2. The level of primary tone f_2 (L_2) ranged from 30 to 50 dB SPL in 5-dB steps, and the level of primary tone f_1 was set equal to $L_1 = 0.4L_2 + 39$, the rule proposed by Kummer et al. (1998) to obtain largest DPOAEs for $L_2 \leq 65$ dB SPL. The duration of the primary tones was 225 ms, and the inter-tone duration was 42 ms. A DPOAE measure for a given f_2 and L_2 combination included 10 stimulus trials. Eight DPOAE measures (of 10 trials each) were obtained for each f_2 and L_2 combination, i.e., four measures were obtained with CAS and four measures were obtained without CAS in interleaved order. The criteria used to calculate the mean DPOAE level across measures were the same as for CEOAEs.

The $2f_1 - f_2$ DPOAE recorded in the ear canal is the vector sum of an OAE distortion component generated at the cochlear region tuned around the f_2 primary tone and an OAE reflection component generated at the $2f_1 - f_2$ cochlear region (Shera and Guinan, 1999; Talmadge et al., 1999; Kalluri and Shera, 2001; Shera and Abdala, 2012). CAS can affect the distortion and reflection components differently, and thus cause DPOAE levels to be sometimes greater in the CAS than in the control condition (Abdala et al., 2009; Deeter et al., 2009; Henin et al., 2011). For this reason, a suppressor tone near the $2f_1 - f_2$ frequency was used in an attempt to suppress the reflection-source contribution to DPOAE (Heitmann et al., 1998; Talmadge et al., 1999; Kalluri and Shera, 2001; Konrad-Martin et al., 2001; Johnson et al., 2006). The suppressor frequency was 64, 59, 54, 54, and 54 Hz below $2f_1 - f_2$ when f_2 was 1, 1.5, 2, 3, or 4 kHz, respectively. The levels of the suppressor (L_3) were calculated according to Figure 8 of Johnson et al. (2006). However, because Johnson et al. (2006) observed variability of up to 15 dB in the optimal suppressor level across participants, we decided to use a suppressor level 10 dB below the level determined by the linear fit for their group data. We made that decision in an attempt not to affect the distortion-source component for subjects who needed a lower suppressor level than the mean to remove the reflection-source component contribution. We used the data centered at 2 kHz from Johnson et al. (2006) to calculate the suppressor levels for $f_2 = 1$, 1.5, and 2 kHz, and the data centered at 4 kHz to calculate the suppressor levels for $f_2 = 3$ and 4 kHz.

The CAS had the same characteristics as described for tone detection thresholds with the exception of the duration. Here, the CAS onset and offset was controlled manually by the experimenter, as for CEOAEs.

Apparatus

Pure tones and CAS were generated with custom-made Matlab software and played via an RME Fireface 400 soundcard at a sampling rate of 44.1 kHz, and with 24-bit resolution. Stimuli were presented to the participants using Etymotic ER-2 insert earphones. These earphones are designed to give a flat frequency response at the eardrum and have a nominal inter-aural attenuation of 70 dB that minimizes cross hearing. Stimuli were calibrated by coupling the earphones to a sound level meter (Brüel and Kjaer 2238) through a Zwislocki coupler (Knowles DB-100). Calibration was performed at 1 kHz and the measured sensitivity was applied to all frequencies.

CEOAE and DPOAE were measured using an Intelligent Hearing Systems Smart device (with SmartOAE software version 5.10) equipped with an Etymotic ER-10D probe. CEOAE stimuli were calibrated with a Zwislocki coupler (Knowles DB-100) by measuring peak intensity with a sound level meter (Brüel and Kjaer 2238). DPOAE stimuli were calibrated with the same Zwislocki coupler for each primary frequency (f_1 and f_2). In-the-ear pressure calibration was not performed. The system artifact was assessed by presenting clicks at different levels (CEOAEs) and different combinations of primary frequencies and levels (DPOAEs) to a microphone connected to the coupler.

Participants sat in a double-wall sound attenuating booth during all measurements. For tone detection thresholds, earphones were removed between each pair of measurements with and without CAS. Threshold pairs for a given probe frequency could be measured in the same or in different sessions, depending on the availability of the participant. The time lapse between sessions ranged from minutes to a few days (2.2 days on average). During OAE measurements, participants were asked to remain as steady as possible. The OAE probe remained in the participant's ear throughout the whole OAE measurement session to minimize measurement variance from

[1]CEOAEs were measured only over a 9-dB range because 51 dB pSPL was the lowest level at which participants showed valid OAEs (i.e., very few participants showed valid responses at lower levels) and 60 dB pSPL was the highest level we could use without large artifacts in our system using the linear mode of stimulation.

altering the position of the probe in the ear canal. During OAE measurements, we did not control if participants were attending to the stimulus.

Quantification of CAS Effects

Contralateral acoustic stimulation was expected to activate the contralateral MOCR, and thus to linearize BM I/O curves by inhibiting the gain of the BM at low-to-moderate levels (**Figure 1A**). Assuming that the BM response at the tone detection threshold is the same with and without CAS, we expected tone detection thresholds to be higher (worse) with than without CAS (**Figure 1A**). Because DPOAEs and CEOAEs require OHC-mediated amplification and CAS reduces such amplification, we also expected CAS to suppress CEOAEs and DPOAEs (**Figure 1B**). We quantified the CAS effect as the difference (in dB)[2] in tone threshold, CEOAE level and DPOAE level in the CAS minus the control (no-CAS) condition, such that a positive threshold difference or a negative OAE difference would be consistent with BM inhibition/linearization.

It would not be appropriate, however, to compare the increase in tone detection threshold with the suppression of OAE levels

[2]Other authors (e.g., Mishra and Lutman, 2013) have quantified the effect of CAS in percentage as follows: 100 × (OAE_{CTR}-OAE_{CAS})/OAE_{CTR}, with all OAE quantities in units of Pascal. We opted to calculate CAS suppression in dB instead of percentage because the two methods would lead to the same conclusions. This is because a given amount of suppression in dB always corresponds with the same amount of suppression in percentage, independent of the baseline (without CAS) response.

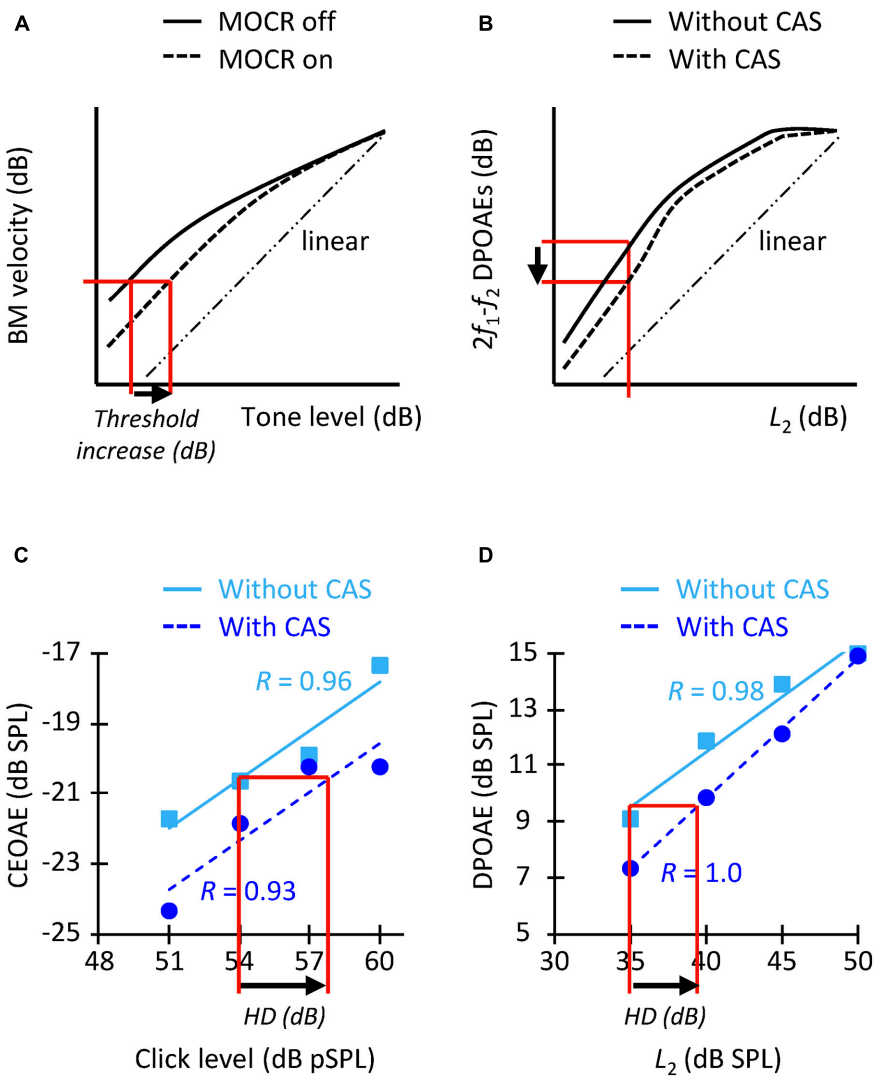

FIGURE 1 | (A) Schematic representation of how MOCR activation is expected to change BM I/O curves (adapted from Cooper and Guinan, 2006) and increase tone detection thresholds (arrow). **(B)** Representation of how MOCR activation is expected to change DPOAE I/O curves [adapted from Moulin et al. (1993)] and reduce the DPOAE level for a given L_2 level (arrow). **(C)** CEOAE I/O functions without (squares) and with (circles) CAS for an example participant. HD indicates the horizontal displacement (in dB). *R*-values indicate the correlation between the data and the linear fit. The I/O functions are for the frequency band of 1.5 kHz. **(D)** DPOAE I/O functions for f_2 = 1.5 kHz for the same participant as in panel **(C)**. The layout is similar as for panel **(C)**.

at any one click level or L_2 because the former presumably quantifies the horizontal displacement of BM I/O curve (also termed "effective attenuation"; Puria et al., 1996; Lichtenhan et al., 2016) (**Figure 1A**) while the latter probably quantifies the vertical displacement of the curve (**Figure 1B**). (Note that the horizontal and vertical displacements of the BM I/O curve are different when responses fall within the compressive region of the I/O curve). For this reason, we also quantified the CAS effect as the horizontal displacement of the CEOAE and DPOAE I/O curves. To do it, we first fitted straight lines to the data without and with CAS (**Figures 1C,D**). The fitting was done only when CEOAEs were present for at least two of the four click levels and when DPOAEs were present for at least three of the five L_2 levels. The correlation between the fit and the data was $\geq$0.90 for 86% of I/O curves with more than three data points, which shows that the choice of a linear fit was appropriate. An I/O curve was excluded from the analyses when the correlation of the fit was <0.75 in the condition with or without CAS (7% of the cases). The horizontal displacement of CEOAE I/O curves was then calculated by estimating the CEOAE level in the fitted line without CAS produced by a click of 54 dB pSPL, followed by the click level in the CAS-fitted function that produced that same CEOAE level. The horizontal displacement was the difference between this latter value and 54 dB pSPL (arrow in **Figure 1C**). The CEOAE level without CAS for 54 dB pSPL clicks was obtained by extrapolation when the subject had valid CEOAE responses for two higher click levels. The same procedure was applied to estimate the horizontal displacement of DPOAE I/O curves, except that the displacement was calculated relative to the DPOAE responses for $L_2 = 35$ dB SPL (arrow in **Figure 1D**). We calculated the shifts relative to 54 dB pSPL clicks and $L_2 = 35$ dB SPL because very few participants had OAEs at lower click and L_2 levels (**Table 1**).

Quantification of the Reliability of the CAS Effect

The test–retest variability of the different estimates of MOCR strength was assessed in two ways. First, we correlated the magnitude of the CAS effect for trials #1, #2, and #3 and fitted a straight line to the data. If the measures were reliable, i.e., if CAS effect were equal across the three repetitions, the slope would be equal to 1.

The second analysis involved calculating the standard deviation of the CAS effect across trials #1, #2, #3, and/or #4. The more reliable measure would produce the smallest SD across trials. For tone detection thresholds, the SD of the CAS effect was calculated for the three measures at a given frequency. (Fourth measures were not included in the analysis because they were obtained only for some participants and frequencies; see above). For CEOAEs and DPOAEs, the CAS effect was calculated for the first, second, third, and fourth measures, and then the SD of the CAS effect across these measures was calculated. The SD of the CAS effect was calculated when at least three of the four pair of measures (with and without CAS) were valid. In this case, we did not request the across-measures SD to be $\leq$3 dB because enforcing that criterion would have reduced the actual SD.

Statistical Analyses

Statistical analyses were performed using IBM SPSS v. 23. Normality was tested with the Shapiro–Wilk test, and parametric

TABLE 1 | Statistical significance of the CAS effect for the different conditions of the study.

			CEOAEs				DPOAEs				
Freq. (kHz)		**Tone thresholds**	**Click = 51 dB pSPL**	**Click = 54 dB pSPL**	**Click = 57 dB pSPL**	**Click = 60 dB pSPL**	**L_2 = 30 dB SPL**	**L_2 = 35 dB SPL**	**L_2 = 40 dB SPL**	**L_2 = 45 dB SPL**	**L_2 = 50 dB SPL**
Overall	*N*		5	12	8	8					
	p		1.000†	1.000†	0.558†	1.000					
0.5	*N*	28									
	p	**<0.001**									
1	*N*		7	17	15	15	4	8	11	12	11
	p		0.395	0.114†	0.095	**0.019**	1.000	0.345†	1.000†	0.115†	0.147
1.5	*N*	28	7	18	16	17	7	14	14	15	15
	p	**<0.001**	0.546†	**0.005**†	0.066†	**0.004**	0.161	**<0.001**	**<0.001**	**0.008**	**0.003**
2	*N*		5	13	13	15	6	7	10	14	13
	p		1.000	0.652	0.855	**0.024**†	1.000	0.854	0.062	**0.010**†	1.000
3	*N*		10	21	16	17	5	6	12	13	14
	p		**0.045**	**0.005**†	**0.002**	**0.003**	1.000	0.375†	0.266	1.000	0.769
4	*N*	28	7	11	10	13	7	10	12	13	15
	p	**<0.001**	**0.007**	**0.042**	**0.001**	**0.003**	1.000	0.305	1.000	0.664	**0.028**

For a given condition, statistical significance was tested with non-parametric Wilcoxon signed rank test (†) or with paired t-test, as appropriate. p-values are corrected for multiple comparisons. Statistically significant effects ($p \leq 0.05$) are highlighted using bold font. Note that the number of CEOAE and DPOAE responses used in the tests (N) is smaller than the number of participants (28 participants for tone thresholds and 54 dB pSPL clicks; 20 participants for 60 dB pSPL clicks; 18 participants for other conditions). This is because some measurements did not meet the inclusion criteria (see section "Materials and Methods") and precluded us from using RMANOVAs or Friedman's tests. Mean and individual results for each condition are depicted in ***Figure 2****.*

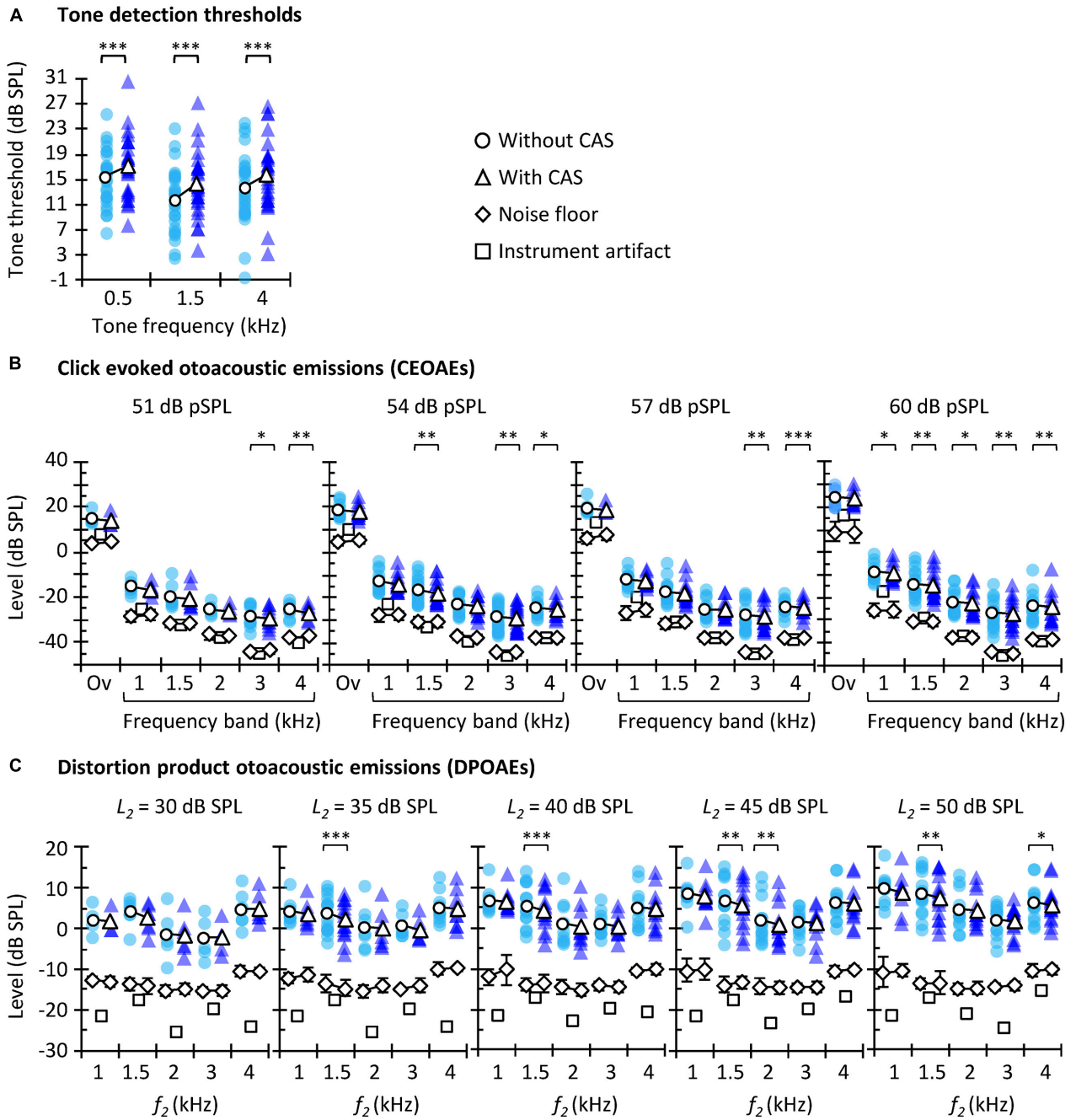

FIGURE 2 | Tone detection thresholds **(A)**, CEOAEs **(B)**, and DPOAEs **(C)** without (circles) and with (triangles) CAS. Open symbols depict mean data and filled symbols depict individual results. For CEOAEs, scores are shown for different click levels and frequency bands, as well as for the overall ("Ov") response, as indicated in each panel. For DPOAEs, scores are shown for the different L_2 and f_2. The background noise level (±one SD) is illustrated by diamonds, both without CAS (diamonds underneath the circles) and with CAS (diamonds underneath the triangles). The instrument artifact level is illustrated by squares. The exact number of participants for each probe frequency and level as well as exact p-values are shown in **Table 1**. Asterisks indicate statistically significant Bonferroni-corrected pairwise comparisons at $^*p \leq 0.05$, $^{**}p \leq 0.01$, $^{***}p \leq 0.001$.

or non-parametric tests were used as appropriate to evaluate the statistical significance of the CAS effect on tone detection thresholds, CEOAEs, and DPOAEs (**Figure 2**), as well as to evaluate the CAS effect for different probe frequencies (**Figure 3**) and levels (**Figure 4**). Pearson's coefficient of correlation was used to investigate if there was a correlation between the different estimates of contralateral MOCR strength (**Figures 5**, **6**). A score was regarded an outlier when it was outside 1.5 times the interquartile range. Outliers were not included in the correlations.

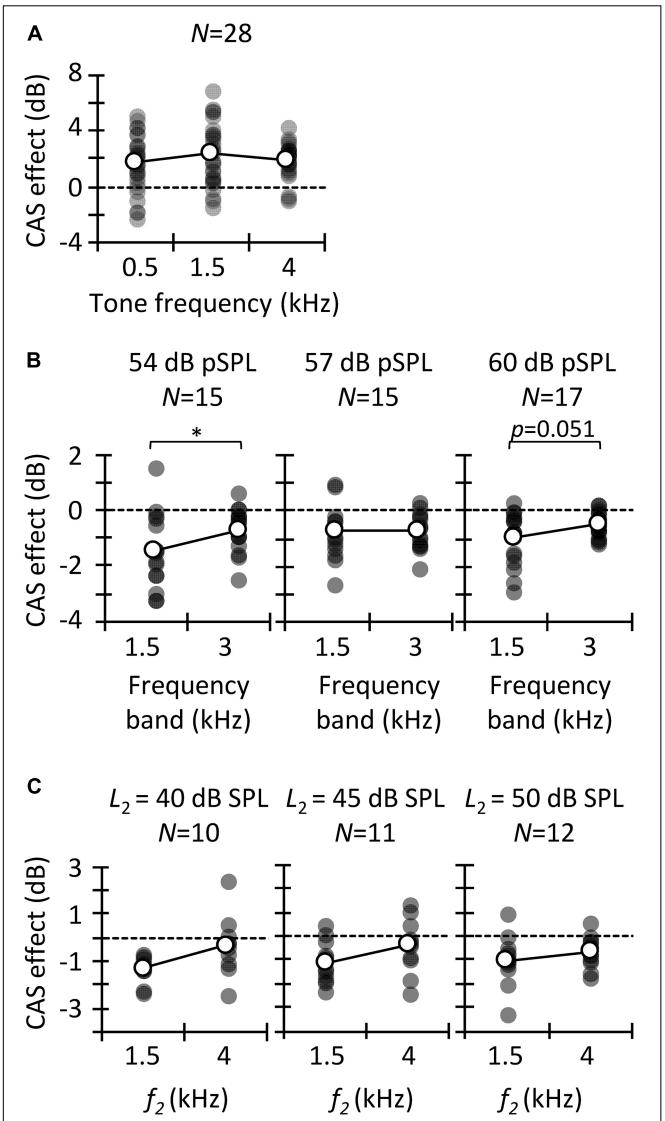

FIGURE 3 | Contralateral acoustic stimulation (CAS) effect for different probe frequencies. Results are for tone detection thresholds **(A)**, CEOAEs **(B)**, and DPOAEs **(C)**. Open symbols depict mean data and filled symbols depict individual results. Data for a given click level or L_2 are for the same participants, as indicated (*N*). Asterisks indicate statistically significant differences in pairwise comparison at $p \leq 0.05$.

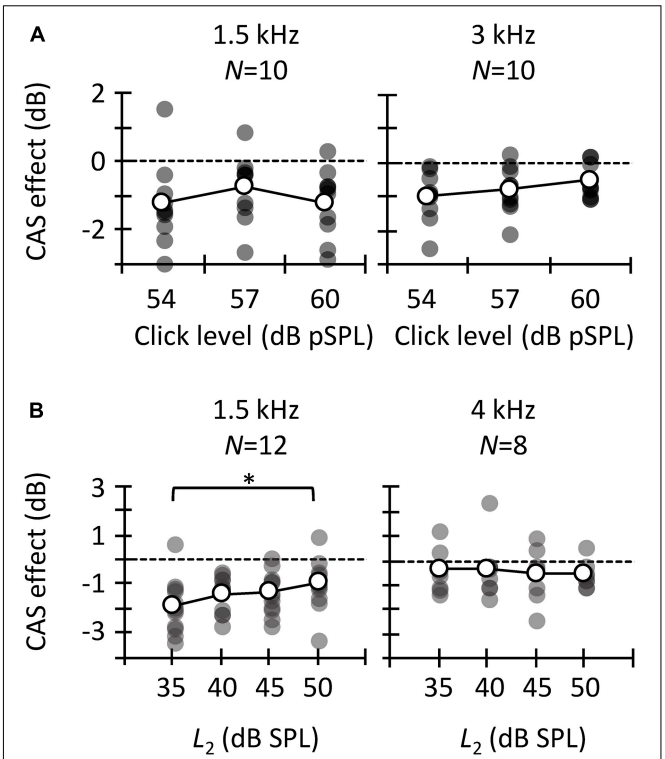

FIGURE 4 | Contralateral acoustic stimulation (CAS) effect as a function of probe level. Results are for CEOAEs **(A)** and DPOAEs **(B)**. Data for a given frequency band **(A)** or f_2 **(B)** are for the same participants (*N*). Open symbols depict mean data and filled symbols depict individual results. Asterisks indicate statistically significant differences in the pairwise comparison at $p \leq 0.05$.

Because OAE data were not available for all participants and conditions (**Table 1**), the statistical tests used in the study focused on optimizing the analyses of the available data. For example, for any given click level, multiple *t* tests instead of a repeated-measures analysis of the variance (RMANOVA) were used to analyze the effect of CAS at every test frequency (**Figure 2B**); the RMANOVA would have excluded participants with missing data in some conditions. Similarly, for CEOAEs and DPOAEs, correlations were performed separately for each probe frequency and level instead of averaging data across all stimulus frequencies and/or levels, something that would have been interesting.

We applied two-tailed tests for all analyses. An effect was regarded as statistically significant when the null hypotheses could be rejected with 95% confidence ($p \leq 0.05$). Unless otherwise stated, we applied Bonferroni corrections for multiple pairwise comparisons.

RESULTS AND DISCUSSION

CAS Effect on Tone Detection Thresholds, CEOAEs, and DPOAEs

The aims of this study are to investigate (1) the correlation between three different methods often used to assess MOCR strength in humans; and (2) which of the three methods is more reliable. Before addressing these aims, however, we explored if the CAS had the expected effect of increasing tone thresholds and suppressing OAEs. **Figure 2** shows tone detection thresholds (**Figure 2A**), CEOAE levels (**Figure 2B**), and DPOAE levels (**Figure 2C**) for all participants and test conditions. Note that there are fewer data points than participants were tested because some data did not meet the inclusion criteria (see section "Materials and Methods"). Average CEOAEs and DPOAEs were 13.9 and 16.6 dB, respectively, above the average noise floor (mean across probe levels and frequencies). These

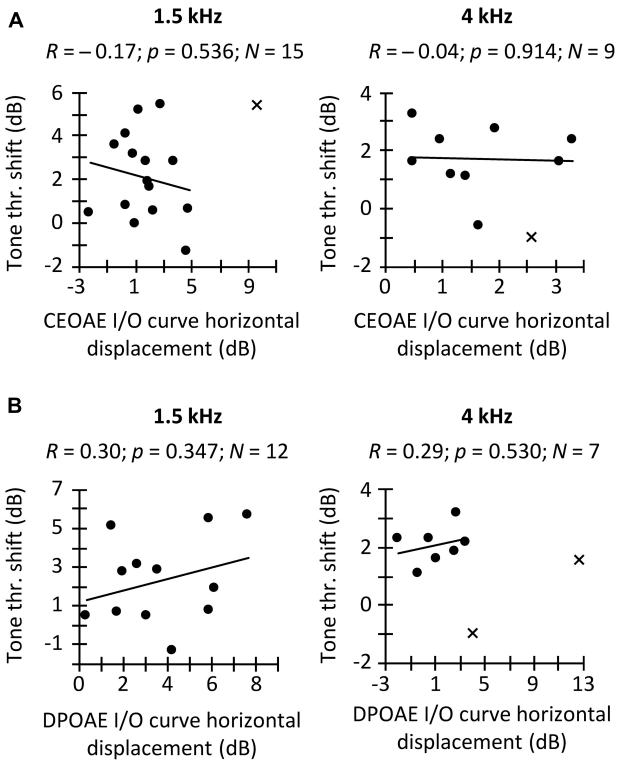

FIGURE 5 | Correlation between the CAS effect on tone detection thresholds and on the horizontal displacement of CEOAE **(A)** and DPOAE **(B)** I/O curves. Separate panels are shown for frequency bands of 1.5 and 4 kHz **(A)** and for f_2 of 1.5 and 4 kHz **(B)**, as indicated at the top of each panel. Data included in the correlation are depicted with circles, and outliers are depicted with crosses. A result was regarded as an outlier when it was outside 1.5 times the interquartile range. *R*-values indicate Pearson correlation coefficient. *p*-values were not corrected for multiple comparisons.

values indicate good quality of the OAEs recorded. Analyses showed that CAS increased tone detection thresholds and tended to suppress CEOAEs and DPOAEs, as expected. This trend occurred for all conditions, but the number of statistically significant pairwise comparisons was relatively greater at higher that at lower probe levels [i.e., for 54 and 60 dB pSPL click levels (**Figure 2B**) or L_2 = 45 or 50 dB SPL (**Figure 2C**)] probably because of the larger number of data points at higher levels (**Table 1**).

CAS Effect as a Function of Probe Frequency

We analyzed the CAS effect as a function of probe frequency to investigate to what extent our results are consistent across the three MOCR estimates as well as with previous studies. **Figure 3A** depicts the CAS effect on tone detection thresholds. The mean (±SD) magnitude of the CAS effect was 1.7 (±2.0), 2.3 (±2.2), and 2.0 (±1.2) dB for 0.5, 1.5, and 4 kHz, respectively. Friedman's test did not reveal significant differences in the magnitude of CAS effect across frequency [$\chi^2(2) = 1.5$, $p = 0.472$]. This result is consistent with Aguilar et al. (2015) and Nogueira et al. (2019), who did not find significant differences in the effect of CAS on detection thresholds for 0.5- and 4-kHz tones when the duration of the tones was ≥200 ms. By contrast, Kawase et al. (2003) found greater CAS effect at 2 kHz than at lower or higher frequencies. On the other hand, the magnitude of the present CAS effect is comparable to that reported elsewhere (Aguilar et al., 2015; but see Kawase et al., 2003).

The magnitude of CAS effect on the 1.5- and 3-kHz components of CEOAEs is depicted in **Figure 3B**. Results are shown only at the two frequencies with the largest number of participants (**Table 1**). For 54 dB pSPL clicks, CEOAE suppression was greater at 1.5 than at 3 kHz (−1.4 vs. −0.7 dB) [$t(14) = -2.25$; $p = 0.041$]. For 57 dB pSPL clicks, CEOAE suppression was similar at 1.5 and 3 kHz (−0.75 vs. −0.72) [$t(14) = -0.10$; $p = 0.921$]. For 60 dB pSPL clicks, CEOAE suppression tended to be greater at 1.5 than at 3 kHz (−1.0 vs. −0.5 dB) [$t(16) = -2.11$; $p = 0.051$]. Our results are consistent with previous studies that found greater CAS effect on CEOAEs for frequency bands centered at or around 1.5 kHz than at 3 kHz (Francis and Guinan, 2010; Lisowska et al., 2014). In addition, the magnitude of CEOAE level suppression is in line with previous studies. For example, Francis and Guinan (2010) used 50 dB pSPL clicks and a contralateral BBN of 60 dB SPL and found CEOAE suppression about −1.5 dB for frequency bands ≤ 2.75 kHz, and about −0.5 dB for frequency bands between 3.25 and 5.25 kHz. Those values are close to the present estimates for 54 dB pSPL clicks, the closest level.

The magnitude of CAS effect on DPOAEs is depicted in **Figure 3C** for test frequencies of 1.5 and 4 kHz. Paired *t*-tests did not reveal a significant probe frequency effect when L_2 was 40 dB SPL [$t(9) = -1.98$; $p = 0.079$], 45 dB SPL [$t(10) = -1.54$; $p = 0.155$], or 50 dB SPL [$t(11) = -1.09$ $p = 0.300$], although suppression tended to be greater at 1.5 than at 4 kHz. The mean suppression was −1.3 and −0.3 dB at 1.5 and 4 kHz, respectively, for L_2 = 40 dB SPL; −1.4 and −0.4 dB, respectively, for L_2 = 45 dB SPL; and −1.0 and −0.7, respectively, for L_2 = 50 dB SPL. Previous studies have reported a gradual reduction of CAS effect with increasing primary frequencies (Wagner and Heyd, 2011; Abdala et al., 2014; Lisowska et al., 2014; Wicher and Moore, 2014), consistent with the present trend. The magnitude of DPOAE suppression is also consistent with previous studies. For example, Wicher and Moore (2014) reported an across-frequency mean DPOAE suppression of −1.4 (± 0.8) dB with contralateral BBN of 60 dB SPL and L_2 = 50 dB SPL.

Altogether, the trend and magnitude of present CAS effects are consistent with those reported in previous studies. We found that the CAS effect on CEOAEs and DPOAEs tended to be greater at lower frequencies whereas it was fairly constant across frequencies for tone detection thresholds. This shows that the frequency dependence of the CAS effect was inconsistent for behavioral and OAEs measures.

CAS Effect as a Function of Probe Level

Most physiological studies have shown that MOC activation suppresses BM responses (Murugasu and Russell, 1996; Dolan et al., 1997; Cooper and Guinan, 2006) and the compound

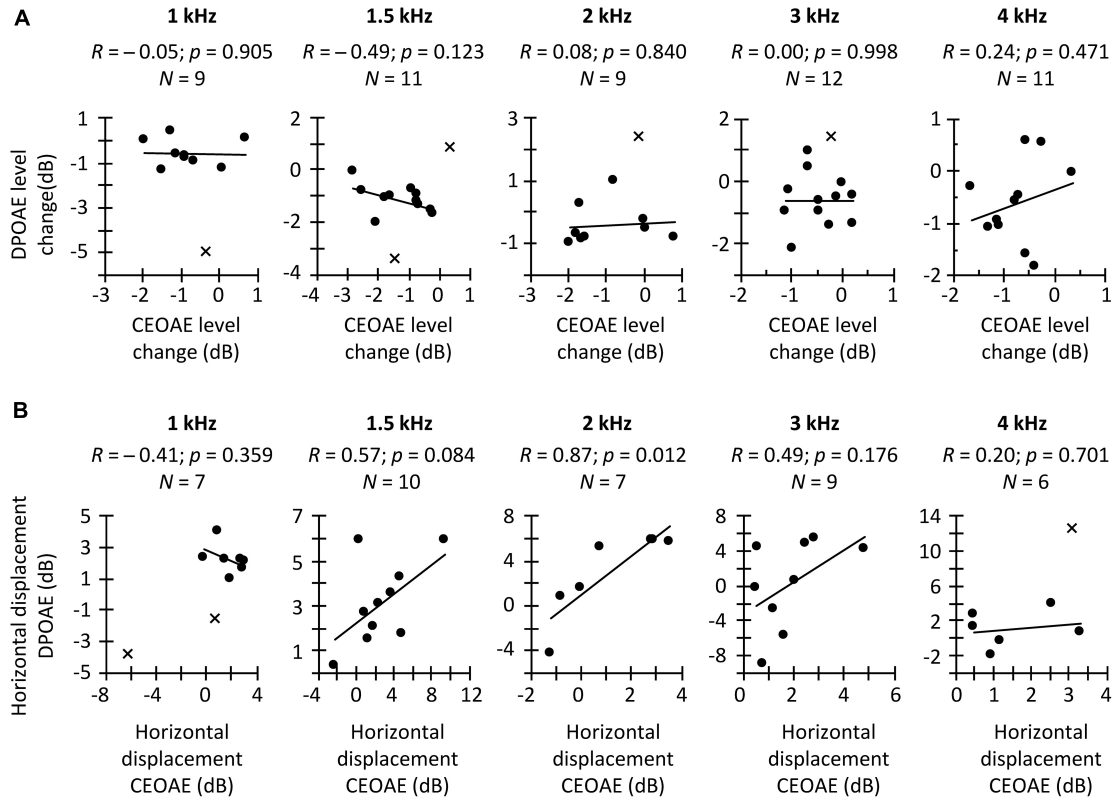

FIGURE 6 | (A) Correlation between the CAS effect on CEOAEs for 60 dB pSPL clicks and on DPOAEs for $L_2 = 50$ dB SPL. **(B)** Correlation between the CAS-induced horizontal displacement on CEOAE and DPOAE I/O curves. Correlations are for different frequency bands (CEOAE) or f_2 (DPOAE), as indicated in each panel. Circles depict data points included in the correlation, while crosses depict outliers not included in the correlation. An outlier is not shown at 2-kHz in panel **(B)** because it was well outside the range of valid results. *R*-values indicate Pearson correlation coefficient. *p*-values were not corrected for multiple comparisons.

action potential (Puria et al., 1996) more at lower than at higher levels, that is, over the range of stimulus levels where the cochlear amplifier gain is greatest (Robles and Ruggero, 2001). The CAS-induced suppression of CEOAEs is also usually greater at lower than at higher click levels (Hood et al., 1996; Veuillet et al., 1996; De Boer and Thornton, 2007; De Boer et al., 2012; Mishra and Lutman, 2013). **Figure 4A** shows the CAS effect on CEOAE levels as a function of click level for the 1.5 and 3 kHz frequency bands. The amount of CEOAE suppression was not significantly different for 54, 57, and 60 dB pSPL clicks neither at 1.5 kHz [one-way RMANOVA: $F(2,18) = 1.46$, $p = 0.258$] nor at 3 kHz [one-way RMANOVA: $F(2,18) = 2.17$, $p = 0.143$]. The absence of a level effect may be due to the narrow range of click levels studied. For example, Hood et al. (1996) found CEOAE suppression to be similar for 50, 55, and 60 dB pSPL clicks, and greater for those lower levels than for 65 or 70 dB pSPL clicks.

CAS-induced DPOAE suppression is also usually greater for lower than for higher primary levels (Moulin et al., 1993; Abdala et al., 1999; Wagner and Heyd, 2011). We found a statistically significant effect of primary level on DPOAE suppression at 1.5 kHz [one-way RMANOVA: $F(3,33) = 2.91$, $p = 0.049$] (**Figure 4B**). Pairwise comparisons with Bonferroni corrections showed greater DPOAE suppression for $L_2 = 35$ dB SPL than for $L_2 = 50$ dB SPL ($p = 0.036$). By contrast, we did not find a statistically significant effect of L_2 level at 4 kHz [Friedman test: $\chi^2(3) = 0.60$, $p = 0.896$].

Within-Subject Correlation of CAS Effect Across Methods

In a first analysis, we investigated the hypothesized within-subject correlation between CAS-induced increase in tone detection threshold and the horizontal displacement of CEOAE or DPOAE I/O curves. Results are shown in **Figure 5**. We found the expected trend only for DPOAEs, although the correlations were far from statistically significant (**Figure 5B**). The pattern of trends suggests that increasing the sample size might bring the correlation between threshold shifts and the horizontal displacement of DPOAE I/O curves closer to statistical significance but would unlikely reveal a correlation between threshold shifts and CEOAE I/O curve shifts. In other words, although the CAS-induced increase in tone detection thresholds and the horizontal displacement of CEOAEs I/O curves are both expected to be the result of the MOCR linearizing BM responses (**Figure 1**), those measures are not equivalent in revealing MOCR effects, at least when using the limited range of click levels used here. Our result agrees with Fletcher et al. (2016), who did

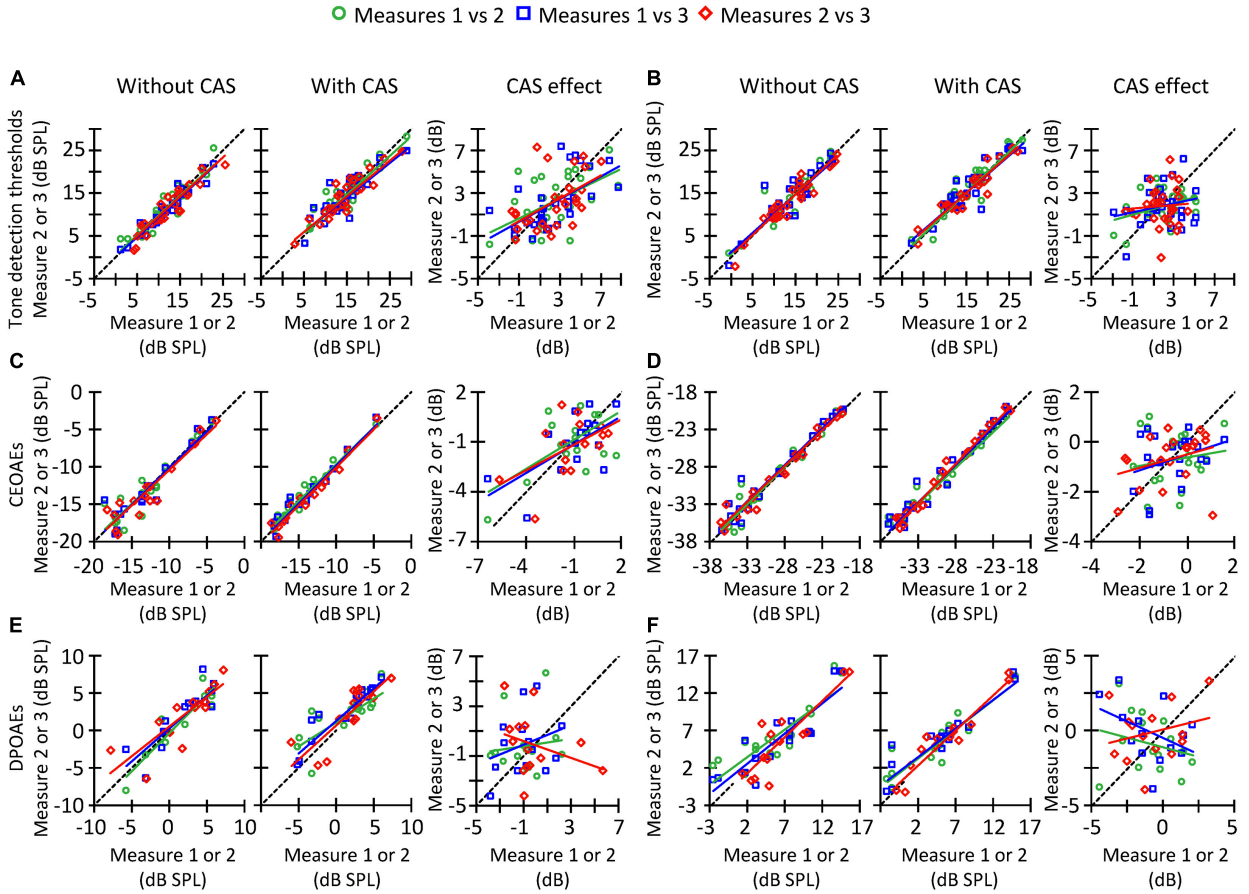

FIGURE 7 | Correlation between trials #1 and #2 (circles), #1 and #3 (squares), or #2 and #3 (diamonds) of tone detection thresholds, CEOAEs, and DPOAEs in the conditions without CAS (left panels), with CAS (mid panels), as well as for the CAS effect (right panels). **(A)** Tone detection thresholds at 1.5 kHz. **(B)** Tone detection thresholds at 4 kHz. **(C)** CEOAEs for 54 dB pSPL clicks and for the frequency band of 1 kHz. **(D)** CEOAEs for 54 dB pSPL clicks and for the frequency band of 3 kHz. **(E)** DPOAEs for $L_2 = 50$ dB SPL and $f_2 = 3$ kHz. **(F)** DPOAEs for $L_2 = 50$ dB SPL and $f_2 = 4$ kHz. The dashed line indicates a 1-to-1 relation across trials, thus the points located on that line indicate highly reliable measures. Each data point represents the results for one participant.

not find a correlation between CEOAE suppression and the reduction of cochlear mechanical gain inferred from temporal masking curves.

In a second analysis, we investigated the hypothesized within-subject correlation between the CAS-induced suppression of CEOAEs and DPOAEs for CEOAEs and DPOAEs obtained at fixed, and roughly matched levels. BM responses to tones can be predicted from BM responses to clicks (Recio et al., 1998) but click and tone levels must be different to obtain the same BM response magnitude with the two stimuli. For example, in Recio et al. (1998), BM responses to 54-dB pSPL clicks predicted accurately the magnitudes of BM responses to 40-dB SPL tones in the chinchilla cochlea. Here, it is hard to know which click level and L_2 produced the same BM response magnitude without CAS. For this reason, we opted to correlate the CAS effect for conditions with the greater number of data points. **Figure 6A** shows the within-subject correlation of the CAS effect on CEOAEs for 60 dB pSPL clicks and on DPOAEs for $L_2 = 50$ dB SPL. The correlation was not significant at any probe frequency. Moreover, the expected trend occurred only at 4 kHz. Complementary analyses (not shown) revealed no significant correlations when using 60 dB pSPL clicks and L_2 of 45 dB SPL, or 57 dB pSPL clicks and L_2 of 50 or 45 dB SPL. The lack of correlation suggests that it is not appropriate to assume that the CAS effects on CEOAEs for a single click level and DPOAEs for a single L_2 provide related information or can be used equivalently. It remains uncertain, however, whether associations would emerge across a broader parametric range (e.g., for other probe levels or averaging several probe levels).

In a third and last analysis, we investigated a potential within-subject correlation between the CAS-induced horizontal displacement of CEOAE and DPOAE I/O curves. **Figure 6B** shows that the expected trend occurred at intermediate frequencies (1.5, 2, and 3 kHz), and that the correlation was indeed statistically significant at 2 kHz. This suggests that the horizontal displacement of CEOAEs and DPOAEs I/O curves may be used somewhat 'equivalently,' at least at these frequencies. It remains uncertain, however, to what extent these displacements are reflecting MOCR effects. For instance, Lichtenhan et al. (2016) measured the CAS effect on the horizontal displacement of

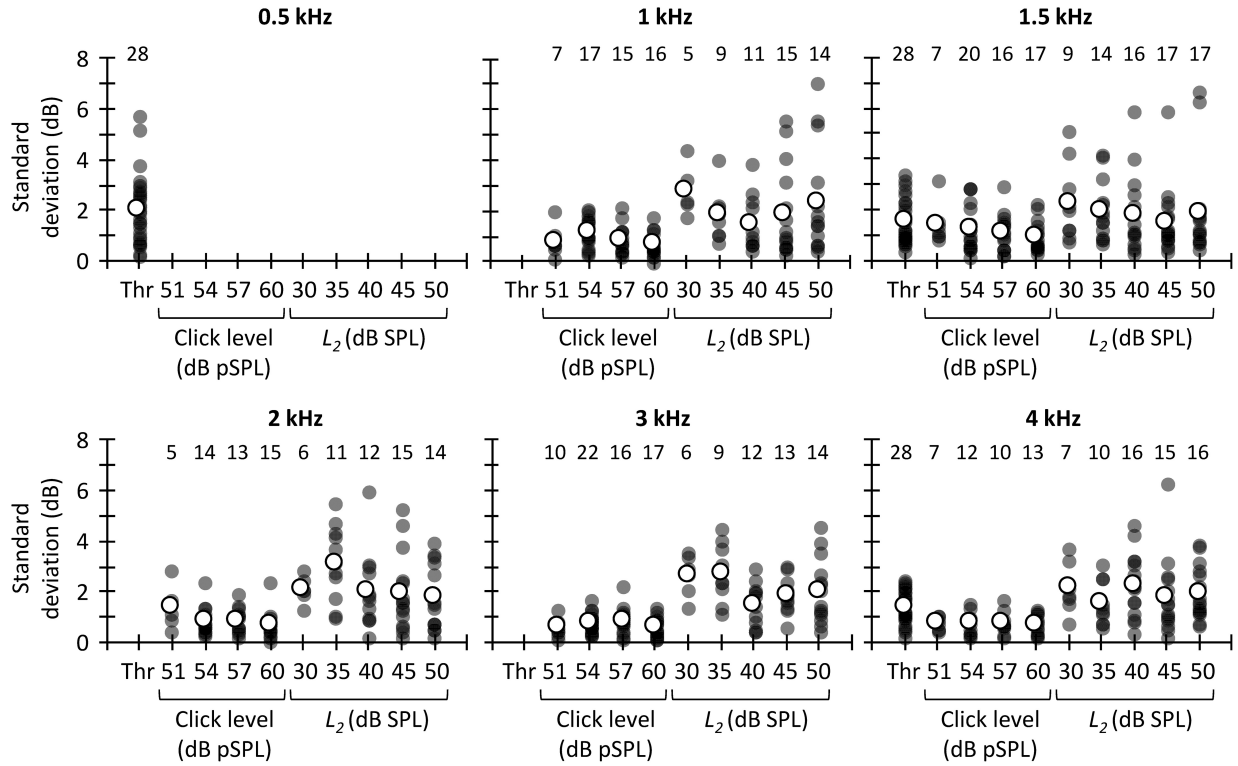

FIGURE 8 | Across-measures standard deviation of CAS effect for tone detection thresholds, CEOAEs, and DPOAEs. Results are for different probe frequencies, click levels (CEOAEs) and L_2 (DPOAEs). Open symbols depict mean data and filled symbols depict individual results. The numbers above each set of circles indicate the number of participants included in the analysis. One outlier for $L_2 = 30$ dB SPL and $f_2 = 3$ kHz is not shown in the figure and was omitted from the mean.

DPOAE I/O functions and on the compound action potential I/O functions in humans and found average trends to be discrepant (their Figure 4), which suggests that factors different from the MOCR are involved in one or both measures.

In summary, the CAS effect on tone detection thresholds was not correlated with the horizontal displacement of CEOAE and DPOAE I/O curves measured in the same subject. Similarly, for fixed stimulus levels, CAS-induced CEOAE suppression was not correlated with CAS-induced DPOAE suppression. The results also showed, however, that the horizontal displacements of CEOAE and DPOAE I/O curves were correlated with each other, at least for mid-frequency probes. The overall lack of correlation can be due to many factors, including the limited parametric space studied (e.g., the clicks and primary levels used here may represent different points in the CEOAE and DPOAE amplitude growth function), the limited resolution of I/O curves (e.g., the 9 dB range of click levels for CEOAE I/O curves may be too narrow to properly define the amplitude growth function), and/or the reduced numbers of observations due to data exclusion. However, other factors such as a low reliability of the measures could be another possible cause (see below).

Reliability of CAS Effects

In the preceding sections, it has been shown that the three different methods used to estimate the MOCR strength are not correlated with each other (**Figures 5**, **6**). This can be partly due to the low reliability of the measures. **Figure 7** illustrates across-trial correlations for tone detection thresholds (**Figures 7A,B**), CEOAEs (**Figures 7C,D**), and DPOAEs (**Figures 7E,F**) in the conditions without (left panels) and with CAS (mid panels), as well as for the CAS effect (right panels). In all panels, the dashed lines illustrate 1-to-1 test–retest correspondence, i.e., zero test–retest variability. For measures obtained with and without CAS, most symbols are located along the dashed line, indicating small test–retest variability. By contrast, for the CAS effect, symbols are away from the dashed line (right-most column in **Figure 7**), indicating high test–retest variability. This variability can be quantified by the slope of a linear fit to the data in each panel of **Figure 7**. For tone detection thresholds at 4 kHz (**Figure 7B**), the slope of the fitted function for measures #2 and #3 (red line) was 0.94 dB/dB in the condition without CAS, 0.88 dB/dB in the condition with CAS, and 0.10 dB/dB for the CAS effect. Because the slope was very different from 1 dB/dB in the latter case, we conclude that the CAS effect on 4-kHz tone detection thresholds is not reliable. Similar patterns are observed for CEOAEs (**Figures 7C,D**) and DPOAEs (**Figures 7E,F**), something surprising considering that OAE measures with and without CAS were obtained without removing the OAE probe from the participant's ear. Altogether, the present results indicate that neither CAS-induced increases in tone threshold nor OAE suppression are reliable estimates of MOCR strength.

If one of the three methods considered here must be chosen to estimate MOCR strength, however, it would be useful to know which one is the most reliable to prioritize its use over the other one(s). **Figure 8** shows the SD of the CAS effect across different trials for tone detection thresholds, CEOAEs and DPOAEs. At all frequencies, the SD was greater for DPOAEs than for tone detection thresholds or CEOAEs, demonstrating that CEOAEs or tone detection thresholds provide more reliable estimates of CAS effects than do DPOAEs. However, as noted earlier, reliability of the CAS effect on tone detection thresholds or CEOAEs can be also low at some frequencies (**Figure 7**).

The greater test–retest reliability of the CAS effect for CEOAEs than for DPOAEs is consistent with previous studies. Stuart and Cobb (2015) and Mertes and Goodman (2016) used Cronbach's alpha, where a value of 1 indicates perfect reliability, to assess the intra-session test–retest reliability of CAS effect on CEOAES. Stuart and Cobb (2015) reported a Cronbach's alpha greater than 0.8, and Mertes and Goodman (2016) reported a Cronbach's alpha greater than 0.95. Kumar et al. (2012) and Kalaiah et al. (2018), however, reported a mean (across frequency) intra-session Cronbach's alpha for DPOAEs of 0.5 and 0.3, respectively. It is unclear why reliability of the CAS effect was lower for DPOAEs than for CEOAEs. Kumar et al. (2012) and Kalaiah et al. (2018) proposed that attentional status might have contributed to the low test–retest reliability of DPOAEs. We, however, measured CEOAEs and DPOAEs in the same participants, and there is no reason to think that attentional status was more variable when measuring DPOAEs than CEOAEs. On the other hand, one might argue that the large test–retest variability for DPOAEs is related with the dual source (reflection and distortion) generation mechanism. DP-grams, however, are highly stable across measurement sessions (Gaskill and Brown, 1990; Zhang et al., 2007). That is, the amplitude and phase of the reflection and distortion component would not change (or not too much) from one trial to another. Hence, although possible, it is uncertain how the MOCR would affect differently the reflection and/or distortion components from one trial to another.

The present SDs of the CAS effect (**Figure 8**) are also in line with previous studies. Stuart and Cobb (2015) reported an intra-session SD of ∼0.4 dB when they used 60 dB pSPL clicks. We observed a mean (across frequency) SD of 0.7 dB for 60 dB pSPL clicks, the most similar condition. Kumar et al. (2012) quantified the variability of the CAS effect on DPOAEs as the standard error across-measures. They reported a mean (across frequency) standard error of 0.8 dB for $L_2 = 55$ dB SPL. Here, the across frequency mean standard error was 1.0 dB for $L_2 = 50$ dB SPL.

GENERAL DISCUSSION

We have shown that, on average, the use of a contralateral broadband noise increased tone detection thresholds, suppressed CEOAEs and DPOAEs, and horizontally shifted CEOAE and DPOAE I/O curves to higher levels, as expected. However, no correlations were found between the CAS effect on tone detection thresholds and on the horizontal displacement of CEOAEs or DPOAEs I/O curves (**Figure 5**), or between the CAS-induced suppression of CEOAEs at DPOAEs for a given stimulus level (**Figure 6A**). The horizontal displacements of CEOAE and DPOAE I/O curves were, however, correlated with each other, at least for the conditions with the greatest number of subjects (**Figure 6B**). We also found that the CAS effect on tone detection thresholds and CEOAEs showed the lowest test–retest variability, suggesting that their use should be prioritized over the use of DPOAEs.

Possible Factors Responsible for the Lack of Correlation Across the Different Measures

The lack of correlation across the different MOCR strength estimates may be due to various factors. First, the restricted parameters employed could be a possible reason. We correlated CAS effects on CEOAEs and DPOAEs for fixed probe levels that may represent different points in the CEOAE and DPOAE amplitude growth function and thus result in weak or absent correlations. In addition, CEOAE and DPOAE I/O curves comprised only 2 or 3 points for some participants. This limited I/O curves resolution may have been insufficient to accurately define the amplitude growth. Further studies should test whether correlations emerge after exploring a broader parametric range.

Second, the CAS-induced increments in tone detection thresholds may reflect 'central masking' in addition to, or instead of, a linearization of BM responses by contralateral MOCR activation. That is, the CAS could have interacted with the test tone in the central auditory nervous system making tone detection harder, a phenomenon referred to as 'central masking.' Evidence in favor of central masking on tone detection thresholds has been reported previously. Smith et al. (2000) demonstrated that, in macaques, the tone threshold increment with CAS remained to some extent when MOC efferents were sectioned. Marrufo-Pérez et al. (2018) showed that detection thresholds for short (50 ms) tones increased more when the tone and CAS onset coincided (early condition) than when the tone onset was delayed 300 ms from the CAS onset (late condition). Because the time course of MOCR activation is around 300 ms (Backus and Guinan, 2006), one would expect greater threshold increments in the 'late' than in the 'early' condition if the MOCR were the only responsible for the increments, but this was not the case. In addition, several studies have demonstrated that bilateral cochlear implant users show an increase in the detection threshold of a probe signal in the presence of contralateral electric stimulation (Van Hoesel and Clark, 1997; James et al., 2001; Lin et al., 2013; Aronoff et al., 2015; Lee and Aronoff, 2018) despite the electrical stimulation delivered by cochlear implants bypasses OHCs and hence is independent from the MOCR. It is hard to differentiate the contribution of the MOCR from central masking. In addition, it is uncertain why some participants showed lower tone detection thresholds with than without CAS (**Figure 3A**), especially considering that both MOCR activation and central masking should have resulted in higher tone detection thresholds.

Third, the attentional state of the participants during the OAE measurements might have affected the results. Several studies

have reported that auditory or visual selective attention can alter transient evoked OAEs (Froehlich et al., 1993a; De Boer and Thornton, 2007; Garinis et al., 2011; Namasivayam et al., 2015), DPOAEs (Smith et al., 2012; Srinivasan et al., 2012, 2014; Wittekindt et al., 2014), or the compound action potential (Delano et al., 2007), presumably by activation of the MOCR. In the present tone-detection experiment, participants must have attended to both visual (the lights displayed in the computer screen) and acoustic cues (the tones) during the measurements. CEOAEs and DPOAEs, by contrast, were recorded without controlling the attentional state of the participant. Therefore, it is uncertain if and to what extent participants were attending to the acoustic stimuli during OAE measurements. Moreover, some participants slept during OAE recordings, and sleeping can decrease efferent activity (Froehlich et al., 1993b). These factors could be partly responsible for the weak (or lack of) correlation between the CAS effect on tone detection thresholds, CEOAEs and/or DPOAES.

Fourth, the middle-ear muscle reflex (MEMR) could have confounded the results to some extent. We set the level of the contralateral BBN at 60 dB SPL because this level has been often used as MOCR elicitor (e.g., Abdala et al., 1999, 2009, 2014; Wagner et al., 2008; Francis and Guinan, 2010; Wicher and Moore, 2014; Aguilar et al., 2015; Boothalingam and Purcell, 2015; Mertes and Leek, 2016). Using the same level of CAS for all participants, however, may not be ideal because some listeners can have a MEMR threshold as low as 50 dB SPL for BBN (Zhao and Dhar, 2010; Feeney et al., 2017). The contraction of the middle-ear muscle changes middle-ear transmission and hence OAEs. If our contralateral stimulation activated the MEMR in some participants but not in others, this would introduce uncertainty and variability in the measures of MOCR strength. In addition, if the probability of MEMR activation was different for DPOAEs, CEOAEs, or threshold measurements, this could have contributed to the poor correlation and reliability of the measures.

Fifth, the lack of correlation between CAS-induced CEOAE and DPOAE suppression for a given stimulus level (i.e., for a given click level and L_2) may have occurred to some uncertain extent because the third tone used when measuring DPOAEs did not suppress totally the reflection component. CAS changes the phase of the reflection component but not (or not so much) the phase of the distortion component, thus resulting in an increase of the DPOAE level when the two components change from canceling each other in the condition without CAS to combining in a constructive fashion in the condition with CAS (Deeter et al., 2009; Henin et al., 2011). As described in the Section "Materials and Methods," the level of the suppressor tone was 10 dB below that suggested by Johnson et al. (2006). This level may have been insufficient to suppress the reflection-source contribution for some participants, which could explain why CAS sometimes enhanced rather suppressed DPOAEs (e.g., **Figures 3C**, **4B**), thus resulting in CAS effect on DPOAEs to be an unreliable MOCR estimate.

Sixth, we did not control for the effects of standing waves in the ear canal, which can be present above 2–3 kHz and lead to inaccurate measurement of stimulus levels. Standing waves occurs when the stimulus presented to the ear canal (forward waveform) interacts with the stimulus reflected from the eardrum (backward waveform). These waveforms can enhance or cancel each other when are in phase or out of phase, respectively, resulting in a difference in the probe level between the microphone and eardrum of up to 20 dB (Stinson et al., 1982; Siegel, 1994). If standing waves were present during OAE recordings, they probably introduced noise into the measurements and consequently, the MOCR gauged by OAEs.

Seventh, as described previously, test–retest repeatability of the CAS effect for tone detection thresholds, CEOAES, and DPOAEs was very low for some probe frequencies (**Figure 7**) despite the various OAE trials (with and without CAS) were measured in a single session without removing the OAE probe. This low reliability can also contribute to the low (or lack of) correlation across the three methods used to estimate the MOCR strength. It is uncertain why the test–retest repeatability was low. One or more of the factors described in the preceding paragraphs (e.g., attentional status) could be responsible for it. Complementary Bland–Altman analyses (Bland and Altman, 1999) revealed that there was not a systematic bias of the measures from trial #1 to #3, i.e., the difference of the CAS effect between trials 1 and 3 was zero on average for tone detection thresholds, CEOAEs, and DPOAEs (results not shown). This indicates that the factor(s) that causes the low repeatability of the measures was independent of trial order.

Eighth, all participants were normal-hearers with presumably normal efferent system reflexes. It is possible that the natural scatter or variation in the MOCR strength was not large enough to capture a correlation well, if it exists.

Lastly, a potential problem is that CAS-induced changes in OAEs level need not reflect the reduction in the cochlear-amplifier gain, as is usually assumed. For example, Berezina-Greene and Guinan (2017) demonstrated that SFOAE amplitude increased at some frequencies and decreased at others when MOC efferents were activated by brainstem shocks in guinea pigs, and the increments occurred despite the animals showed a *reduction* in the cochlear-amplifier gain. Similar results might occur for CEOAEs insofar as CEOAEs and SFOAEs are generated by the same mechanism (Kalluri and Shera, 2007; Francis and Guinan, 2010; Shera and Abdala, 2012). Indeed, we found that CAS sometimes increased CEOAEs in some conditions (e.g., **Figure 3A**).

In summary, because the correspondence between the two OAE indices was explored across a limited (and maybe not always matching) range of stimulus levels and frequencies, and because many factors were not or could not be (e.g., central masking) controlled for, it is not surprising that correlations were not observed. Further research is needed to investigate which factors are mostly responsible for this lack of correlation and how their effects can be controlled for, as well as to design better measures of MOCR strength in humans.

Implications

Current evidence supporting the roles of the MOCR in human hearing is mixed (reviewed by Fuente, 2015; Smith and Keil, 2015; Lopez-Poveda, 2018). The discrepant results across studies

can be due to some extent to the methodology used. As in our study, many previous studies did not control for the attentional state of the participants (e.g., Kumar and Vanaja, 2004; Kim et al., 2006; Bidelman and Bhagat, 2015), the presence of fine structure in DPOAEs (e.g., Kim et al., 2006; Mukari and Mamat, 2008), or the presence of standing waves (e.g., Giraud et al., 1997; Kumar and Vanaja, 2004; Kim et al., 2006; Mukari and Mamat, 2008; Bidelman and Bhagat, 2015). In addition, many studies correlated a single estimate of the contralateral MOCR strength (e.g., CAS-induced suppression of CEOAEs for a single click level or DPOAEs for a single combination of primary levels L_1 and L_2) with performance scores in a psychoacoustical task of interest (e.g., Kumar and Vanaja, 2004; Kim et al., 2006; Mukari and Mamat, 2008; Stuart and Butler, 2012; Mishra and Lutman, 2014; Bidelman and Bhagat, 2015; Mertes et al., 2018, 2019). Here, we have shown that CEOAE suppression for a given click level is not correlated with DPOAE suppression for a given L_2 (**Figure 6A**). Hence, it is not surprising that studies reached different conclusions about the roles of the MOCR in human hearing when the MOCR strength was estimated with two different methods and a single probe level [e.g., Mukari and Mamat (2008) and Bidelman and Bhagat (2015)]. On the other hand, some studies have measured DPOAE or CEOAE suppression by performing only one measure without CAS and another measure with CAS (e.g., Kumar and Vanaja, 2004; Stuart and Butler, 2012; Bidelman and Bhagat, 2015; Mertes et al., 2019). Here, we have shown that the suppression of CEOAEs and DPOAEs can be highly variable from trial to trial (**Figures 7**, **8**). Therefore, it is also not surprising that findings were also discrepant across studies that aimed at investigating the roles of the MOCR in human hearing using the same methodology but assessing MOCR strength with only one measure without and with CAS [e.g., Stuart and Butler (2012) and Bidelman and Bhagat (2015)].

Altogether, our study suggests that many confound factors enter into MOCR measurement and that previous studies may have used a simplistic way of evaluating MOCR strength. How to optimize MOCR measurements must be addressed in further studies. In addition, other ways of analyzing OAEs could be explored. For example, Dragicevic et al. (2019) studied low-frequency (1–35 Hz) oscillatory amplitude changes in DPOAEs and electroencephalography to assess whether cortical oscillations modulate cochlear responses during selective attention. Their results were according to their hypothesis, and they propose the auditory efferent system as the most probable neural pathway responsible for modulating cochlear responses. It is possible that using such novel methods for OAE analyses help to investigate the roles of the MOCR in human hearing.

CONCLUSION

(1) On average, contralateral acoustic stimulation (CAS) increased tone detection thresholds and decreased CEOAE and DPOAE levels in normal hearing listeners.
(2) The magnitude of CAS effect tended to be greater for lower (1.5 kHz) than for higher (3–4 kHz) frequencies for CEOAEs and DPOAEs. The effect of CAS on tone detection thresholds, however, was similar in magnitude for 0.5, 1.5, and 4 kHz probe tones.
(3) The CAS effect on CEOAEs was not different for 54, 57, and 60 dB pSPL clicks. The CAS effect on DPOAEs was greater for $L_2 = 35$ dB SPL than for $L_2 = 50$ dB SPL at 1.5 but not at 4 kHz.
(4) The CAS-induced change on tone detection thresholds was not correlated with the CAS-induced horizontal displacement of CEOAE or DPOAE I/O curves.
(5) The CAS effect on CEOAEs for a given click level was not correlated with the CAS effect on DPOAEs for a given L_2.
(6) The horizontal displacements of CEOAEs and DPOAEs I/O curves induced by CAS tended to be correlated with each other, at least for conditions with the greater number of data points.
(7) The test–retest variability of the CAS effect was high overall but smaller for tone detection thresholds and CEOAEs than for DPOAEs.
(8) The weak correlations and poor reliability observed here could be related with inherent limitations of the study, such as the small range of clicks and L_2 levels used, and/or with factors not related with the MOCR. Nonetheless, the present findings show that the different estimates of the MOCR strength cannot be used independently and assume that they provide similar results.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Comité de Bioética, Universidad de Salamanca. The participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

MM-P performed the research, analyzed the data, and wrote the first draft of the manuscript. MM-P and EL-P edited the manuscript and wrote the manuscript. PJ provided technical tools. EL-P designed the research. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by Ministerio de Ciencia e Innovación (grant PID2019-108985GB-I00) and the European Regional Development Fund.

REFERENCES

Abdala, C., Dhar, S., Ahmadi, M., and Luo, P. (2014). Aging of the medial olivocochlear reflex and associations with speech perception. *J. Acoust. Soc. Am.* 135, 754–765. doi: 10.1121/1.4861841

Abdala, C., Ma, E., and Sininger, Y. S. (1999). Maturation of medial efferent system function in humans. *J. Acoust. Soc. Am.* 105, 2392–2402. doi: 10.1121/1.426844

Abdala, C., Mishra, S. K., and Williams, T. L. (2009). Considering distortion product otoacoustic emission fine structure in measurements of the medial olivocochlear reflex. *J. Acoust. Soc. Am.* 125, 1584–1594. doi: 10.1121/1.3068442

Aguilar, E., Eustaquio-Martin, A., and Lopez-Poveda, E. A. (2013). Contralateral efferent reflex effects on threshold and suprathreshold psychoacoustical tuning curves at low and high frequencies. *J. Assoc. Res. Otolaryngol.* 14, 341–357. doi: 10.1007/s10162-013-0373-4

Aguilar, E., Johannesen, P. T., and Lopez-Poveda, E. A. (2015). Contralateral efferent suppression of human hearing sensitivity. *Front. Syst. Neurosci.* 8:251.

ANSI (1996). *S3.6 Specification for Audiometers*. New York, NY: American National Standards Institute.

Aronoff, J. M., Padilla, M., Fu, Q. J., and Landsberger, D. M. (2015). Contralateral masking in bilateral cochlear implant patients: a model of medial olivocochlear function loss. *PLoS One* 10:e0121591. doi: 10.1371/journal.pone.0121591

Backus, B. C., and Guinan, J. J. (2006). Time-course of the human medial olivocochlear reflex. *J. Acoust. So.c Am.* 119, 2889–2904. doi: 10.1121/1.2169918

Berezina-Greene, M. A., and Guinan, J. J. (2017). Electrically evoked medial olivocochlear efferent effects on stimulus frequency otoacoustic emissions in guinea pigs. *J. Assoc. Res. Otolaryngol.* 18, 153–163. doi: 10.1007/s10162-016-0593-5

Bidelman, G. M., and Bhagat, S. P. (2015). Right-ear advantage drives the link between olivocochlear efferent 'antimasking' and speech-in-noise listening benefits. *NeuroReport* 26, 483–487. doi: 10.1097/wnr.0000000000000376

Bland, J. M., and Altman, D. G. (1999). Measuring agreement in method comparison studies. *Stat. Methods Med. Res.* 8, 135–160. doi: 10.1191/096228099673819272

Boothalingam, S., and Purcell, D. W. (2015). Influence of the stimulus presentation rate on medial olivocochlear system assays. *J. Acoust. Soc. Am.* 137, 724–732. doi: 10.1121/1.4906250

Brown, M. C., de Venecia, R. K., and Guinan, J. J. (2003). Responses of medial olivocochlear neurons. specifying the central pathways of the medial olivocochlear reflex. *Exp. Brain Res.* 153, 491–498. doi: 10.1007/s00221-003-1679-y

Chays, A., Maison, S., Robaglia-Schlupp, A., Cau, P., Broder, L., and Magnan, J. (2003). Are we sectioning the cochlear efferent system during vestibular neurectomy? *Rev. Laryngol. Otol. Rhinol.* 124, 53–58.

Cooper, N. P., and Guinan, J. J. (2003). Separate mechanical processes underlie fast and slow effects of medial olivocochlear efferent activity. *J. Physiol.* 548, 307–312. doi: 10.1113/jphysiol.2003.039081

Cooper, N. P., and Guinan, J. J. (2006). Efferent-mediated control of basilar membrane motion. *J. Physiol.* 576, 49–54. doi: 10.1113/jphysiol.2006.114991

De Boer, J., and Thornton, A. R. D. (2007). Effect of subject task on contralateral suppression of click evoked otoacoustic emissions. *Hear. Res.* 233, 177–123.

De Boer, J., Thornton, A. R. D., and Krumbholz, K. (2012). What is the role of the medial olivocochlear system in speech-in-noise processing? *J. Neurophysiol.* 107, 1301–1312. doi: 10.1152/jn.00222.2011

Deeter, R., Abel, R., Calandruccio, L., and Dhar, S. (2009). Contralateral acoustic stimulation alters the magnitude and phase of distortion product otoacoustic emissions. *J. Acoust. Soc. Am.* 126, 2413–2424. doi: 10.1121/1.3224716

Delano, P. H., Elgueda, D., Hamame, C. M., and Robles, L. (2007). Selective attention to visual stimuli reduces cochlear sensitivity in chinchillas. *J. Neurosci.* 27, 4146–4153. doi: 10.1523/jneurosci.3702-06.2007

Dolan, D. F., Guo, M. H., and Nuttall, A. L. (1997). Frequency-dependent enhancement of basilar membrane velocity during olivocochlear bundle stimulation. *J. Acoust. Soc. Am.* 102, 3587–3596. doi: 10.1121/1.421008

Dragicevic, C. D., Marcenaro, B., Navarrete, M., Robles, L., and Delano, P. H. (2019). Oscillatory infrasonic modulation of the cochlear amplifier by selective attention. *PLoS One* 14:e0208939. doi: 10.1371/journal.pone.0208939

Feeney, M. P., Keefe, D. H., Hunter, L. L., Fitzpatrick, D. F., Garinis, A. C., Putterman, D. B., et al. (2017). Normative wideband reflectance, equivalent admittance at the tympanic membrane, and acoustic stapedius reflex threshold in adults. *Ear. Hear* 38, e142–e160.

Fletcher, M. D., Krumbholz, K., and de Boer, J. (2016). Effect of contralateral medial olivocochlear feedback on perceptual estimates of cochlear gain and compression. *J. Assoc. Res. Otolaryngol.* 17, 559–575. doi: 10.1007/s10162-016-0574-8

Francis, N., and Guinan, J. J. (2010). Acoustic stimulation of human medial olivocochlear efferents reduces stimulus frequency- and click-evoked otoacoustic emission delays: Implications for cochlear filter bandwidths. *Hear Res.* 267, 36–45. doi: 10.1016/j.heares.2010.04.009

Froehlich, P., Collet, L., and Morgon, A. (1993a). Transiently evoked otoacoustic emission amplitudes change with changes of directed attention. *Physiol. Behav.* 53, 679–682. doi: 10.1016/0031-9384(93)90173-d

Froehlich, P., Collet, L., Valatx, J. L., and Morgon, A. (1993b). Sleep and active cochlear micromechanical properties in human subjects. *Hear Res.* 66, 1–7. doi: 10.1016/0378-5955(93)90254-x

Fuente, A. (2015). The olivocochlear system and protection from acoustic trauma: a mini literature review. *Front. Syst. Neurosci.* 9:94.

Garinis, A. C., Glattke, T., and Cone, B. K. (2011). The MOC reflex during active listening to speech. *J. Speech Lang. Hear Res.* 54, 1464–1476. doi: 10.1044/1092-4388(2011/10-0223)

Gaskill, S. A., and Brown, A. M. (1990). The behavior of the acoustic distortion product, 2f1-f2, from the human ear and its relation to auditory sensitivity. *J. Acoust. Soc. Am.* 88, 821–839. doi: 10.1121/1.399732

Giraud, A. L., Garnier, S., Micheyl, C., Lina, G., Chays, A., and Chéry-Croze, S. (1997). Auditory efferents involved in speech-in noise intelligibility. *NeuroReport* 8, 1779–1783. doi: 10.1097/00001756-199705060-00042

Guinan, J. J. (2006). Olivocochlear efferents: Anatomy, physiology, function, and the measurement of efferent effects in humans. *Ear. Hear* 27, 589–607. doi: 10.1097/01.aud.0000240507.83072.e7

Handrock, M., and Zeisberg, J. (1982). The Influence of the efferent system on adaptation, temporary and permanent threshold shift. *Acta Otorhinolaryngol.* 234, 191–195. doi: 10.1007/bf00453630

Heitmann, J., Waldmann, B., Schnitzler, H. U., Plinkert, P. K., and Zenner, H. P. (1998). Suppression of distortion product otoacoustic emissions (DPOAE) near 2f1-f2 removes DP-gram fine structure—Evidence for a secondary generator. *J. Acoust. Soc. Am.* 103, 1527–1531. doi: 10.1121/1.421290

Henin, S., Thompson, S., Abdelrazeq, S., and Long, G. R. (2011). Changes in amplitude and phase of distortion-product otoacoustic emission fine-structure and separated components during efferent activation. *J. Acoust. Soc. Am.* 129, 2068–2079. doi: 10.1121/1.3543945

Hood, L. J., Berlin, C. I., Hurley, A., Cecola, P., and Bell, B. (1996). Contralateral suppression of transient-evoked otoacoustic emissions in humans: intensity effects. *Hear Res.* 101, 113–118. doi: 10.1016/s0378-5955(96)00138-4

James, C., Blamey, P., Shallop, J. K., Incerti, P. V., and Nicholas, A. M. (2001). Contralateral masking in cochlear implant users with residual hearing in the non-implanted ear. *Audiol. Neuro Otol.* 6, 87–97. doi: 10.1159/000046814

Johnson, T. A., Neely, S. T., Kopun, J. G., and Gorga, M. P. (2006). Reducing reflected contributions to ear-canal distortion product otoacoustic emissions in humans. *J. Acoust. Soc. Am.* 119, 3896–3907. doi: 10.1121/1.2200048

Kalaiah, M. K., Lasrado, A., Pinto, N., and Shastri, U. (2018). Short term test-retest reliability of contralateral inhibition of distortion product otoacoustic emissions. *J. Audiol. Otol.* 22, 189–196. doi: 10.7874/jao.2018.00038

Kalluri, R., and Shera, C. A. (2001). Distortion-product source unmixing: A test of the two mechanism model for DPOAE generation. *J. Acoust. Soc. Am.* 109, 622–637. doi: 10.1121/1.1334597

Kalluri, R., and Shera, C. A. (2007). Near equivalence of human click-evoked and stimulus-frequency otoacoustic emissions. *J. Acoust. Soc. Am.* 121, 2097–2110. doi: 10.1121/1.2435981

Kawase, T., Delgutte, B., and Liberman, M. C. (1993). Antimasking effects of the olivocochlear reflex. II. Enhancement of auditory-nerve response to masked tones. *J. Neurophysiol.* 70, 2533–2549. doi: 10.1152/jn.1993.70.6.2533

Kawase, T., Ogura, M., Sato, T., Kobayashi, T., and Suzuki, Y. (2003). Effects of contralateral noise on the measurement of auditory threshold. *Tohoku J. Exp. Med.* 200, 129–135. doi: 10.1620/tjem.200.129

Kemp, D. T., Ryan, S., and Bray, P. (1990). A guide to the effective use of otoacoustic emissions. *Ear. Hear* 11, 93–105. doi: 10.1097/00003446-199004000-00004

Kim, S. H., Frisina, R. D., and Frisina, D. R. (2006). Effects of age on speech understanding in normal hearing listeners: relationship between the auditory efferent system and speech intelligibility in noise. *Speech Commun.* 48, 855–862. doi: 10.1016/j.specom.2006.03.004

Konrad-Martin, D., Neely, S. T., Keefe, D. H., Dorn, P. A., and Gorga, M. P. (2001). Sources of distortion product otoacoustic emissions revealed by suppression experiments and inverse fast Fourier transforms in normal ears. *J. Acoust. Soc. Am.* 109, 2862–2879. doi: 10.1121/1.1370356

Kujawa, S. G., and Liberman, M. C. (1997). Conditioning-related protection from acoustic injury: Effects of chronic deefferentation and sham surgery. *J. Neurophysiol.* 78, 3095–3106. doi: 10.1152/jn.1997.78.6.3095

Kumar, U. A., Methi, R., and Avinash, M. C. (2012). Test/retest repeatability of effect contralateral acoustic stimulation on the magnitudes of distortion product otoacoustic emissions. *Laryngoscope* 123, 463–471. doi: 10.1002/lary.23623

Kumar, U. A., and Vanaja, C. S. (2004). Functioning of olivocochlear bundle and speech perception in noise. *Ear. Hear* 25, 142–146. doi: 10.1097/01.aud.0000120363.56591.e6

Kummer, P., Janssen, T., and Arnold, W. (1998). The level and growth behavior of the 2f1-f2 distortion product otoacoustic emission and its relationship to auditory sensitivity in normal hearing and cochlear hearing loss. *J. Acoust. Soc. Am.* 103, 3431–3444. doi: 10.1121/1.423054

Lee, D. H., and Aronoff, J. M. (2018). Changing stimulation patterns can change the broadness of contralateral masking functions for bilateral cochlear implant users. *Hear Res.* 363, 55–61. doi: 10.1016/j.heares.2018.03.001

Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *J. Acoust. Soc. Am.* 49, 467–677. doi: 10.1121/1.1912375

Liberman, M. C., and Brown, M. C. (1986). Physiology and anatomy of single olivocochlear neurons in the cat. *Hear Res.* 24, 17–36. doi: 10.1016/0378-5955(86)90003-1

Lichtenhan, J. T., Wilson, U. S., Hancock, K. E., and Guinan, J. J. (2016). Medial olivocochlear efferent reflex inhibition of human cochlear nerve responses. *Hear Res.* 333, 216–224. doi: 10.1016/j.heares.2015.09.001

Lilaonitkul, W., and Guinan, J. J. (2009). Human medial olivocochlear reflex: Effects as functions of contralateral, ipsilateral, and bilateral elicitor bandwidths. *J. Assoc. Res. Otolaryngol.* 10, 459–470. doi: 10.1007/s10162-009-0163-1

Lin, P., Lu, T., and Zeng, F. G. (2013). Central masking with bilateral cochlear implants. *J. Acoust. Soc. Am.* 133, 962–969. doi: 10.1121/1.4773262

Lisowska, G., Namyslowski, G., Orecka, B., and Misiolek, M. (2014). Influence of aging on medial olivocochlear system function. *Clin. Interv. Aging* 9, 901–914. doi: 10.2147/cia.s61934

Lopez-Poveda, E. A. (2018). Olivocochlear efferents in animals and humans: from anatomy to clinical relevance. *Front. Neurol.* 9:197.

Maison, S. F., and Liberman, M. C. (2000). Predicting vulnerability to acoustic injury with a noninvasive assay of olivocochlear reflex strength. *J. Neurosci.* 20, 4701–4707. doi: 10.1523/jneurosci.20-12-04701.2000

Maison, S. F., Micheyl, C., Andéol, G., Gallégo, S., and Collet, L. (2000). Activation of medial olivocochlear efferent system in humans: influence of stimulus bandwidth. *Hear Res.* 140, 111–125. doi: 10.1016/s0378-5955(99)00196-3

Maison, S. F., Usubuchi, H., and Liberman, M. C. (2013). Efferent feedback minimizes cochlear neuropathy from moderate noise exposure. *J. Neurosci.* 33, 5542–5552. doi: 10.1523/jneurosci.5027-12.2013

Marrufo-Pérez, M. I., Eustaquio-Martín, A., Bascuas, L. E., and Lopez-Poveda, E. A. (2018). Temporal effects on monaural amplitude-modulation sensitivity in ipsilateral, contralateral and bilateral noise. *J. Assoc. Res. Otolaryngol.* 19, 147–161. doi: 10.1007/s10162-018-0656-x

Mertes, I. B., and Goodman, S. S. (2016). Within- and across-subject variability of repeated measurements of medial olivocochlear-induced changes in transient-evoked otoacoustic emissions. *Ear. Hear* 37, 72–84.

Mertes, I. B., Johnson, K. M., and Dinger, Z. A. (2019). Olivocochlear efferent contributions to speech-in-noise recognition across signal-to-noise ratios. *J. Acoust. Soc. Am.* 145, 1529. doi: 10.1121/1.5094766

Mertes, I. B., and Leek, M. R. (2016). Concurrent measures of contralateral suppression of transient-evoked otoacoustic emissions and of auditory steady-state responses. *J. Acoust. Soc. Am.* 140, 2027–2038. doi: 10.1121/1.4962666

Mertes, I. B., Wilbanks, E. C., and Leek, M. R. (2018). Olivocochlear efferent activity is associated with the slope of the psychometric function of speech recognition in noise. *Ear. Hear* 39, 583–593. doi: 10.1097/aud.0000000000000514

Mishra, S. K., and Lutman, M. E. (2013). Repeatability of click-evoked otoacoustic emission-based medial olivocochlear efferent assay. *Ear. Hear* 34, 789–798. doi: 10.1097/aud.0b013e3182944c04

Mishra, S. K., and Lutman, M. E. (2014). Top-down influences of the medial olivocochlear efferent system in speech perception in noise. *PLoS One* 9:e85756. doi: 10.1371/journal.pone.0085756

Moulin, A., Collet, L., and Duclaux, R. (1993). Contralateral auditory stimulation alters acoustic distortion products in humans. *Hear Res.* 65, 193–210. doi: 10.1016/0378-5955(93)90213-k

Mukari, S. Z., and Mamat, W. H. (2008). Medial olivocochlear functioning and speech perception in noise in older adults. *Audiol. Neurootol.* 13, 328–334. doi: 10.1159/000128978

Murugasu, E., and Russell, I. J. (1996). The effect of efferent stimulation on basilar membrane displacement in the basal turn of the guinea pig cochlea. *J. Neurosci.* 76, 325–332. doi: 10.1523/jneurosci.16-01-00325.1996

Namasivayam, A. K., Wong, W. Y., Sharma, D., and van Lieshout, P. (2015). Visual speech gestures modulate efferent auditory system. *J. Integr. Neurosci.* 14, 73–83. doi: 10.1142/s0219635215500016

Nieder, P., and Nieder, I. (1970a). Stimulation of efferent olivocochlear bundle causes release from low level masking. *Nature* 227, 184–185. doi: 10.1038/227184a0

Nieder, P., and Nieder, I. (1970b). Antimasking effect of crossed olivocochlear bundle stimulation with loud clicks in guinea pig. *Exp. Neurol.* 28, 179–188. doi: 10.1016/0014-4886(70)90172-x

Nogueira, W., Krüger, B., Büchner, A., and Lopez-Poveda, E. A. (2019). Contralateral suppression of human hearing sensitivity in single-sided deaf cochlear implant users. *Hear Res.* 373, 121–129. doi: 10.1016/j.heares.2018.06.001

Puria, S., Guinan, J. J., and Liberman, M. C. (1996). Olivocochlear reflex assays: effects of contralateral sound on compound action potentials versus ear-canal distortion products. *J. Acoust. Soc. Am.* 99, 500–507. doi: 10.1121/1.414508

Recio, A., Rich, N. C., Narayan, S. S., and Ruggero, M. A. (1998). Basilar-membrane responses to clicks at the base of the chinchilla cochlea. *J. Acoust. Soc. Am.* 103, 1972–1989. doi: 10.1121/1.421377

Robles, L., and Ruggero, M. A. (2001). Mechanics of the mammalian cochlea. *Physiol. Rev.* 81, 1305–1352. doi: 10.1152/physrev.2001.81.3.1305

Shera, C. A., and Abdala, C. (2012). "Otoacoustic emissions-Mechanisms and applications," in *Translational Perspectives in Auditory Neuroscience. Hearing Across the Life Span: Assessment and Disorders*, eds K. L. Tremblay and R. F. Burkard (San Diego, CA: Plural Publishing), 123–159.

Shera, C. A., and Guinan, J. J. (1999). Evoked otoacoustic emissions arise by two fundamentally different mechanisms: a taxonomy for mammalian OAEs. *J. Acoust. Soc. Am.* 105, 782–798. doi: 10.1121/1.426948

Siegel, J. H. (1994). Ear-canal standing waves and high-frequency sound calibration using otoacoustic emission probes. *J. Acoust. Soc. Am.* 95, 2589–2597. doi: 10.1121/1.409829

Smith, D. W., Aouad, R. K., and Keil, A. (2012). Cognitive task demands modulate the sensitivity of the human cochlea. *Front. Psychol.* 3:30.

Smith, D. W., and Keil, A. (2015). The biological role of the medial olivocochlear efferents in hearing: separating evolved function from exaptation. *Front. Syst. Neurosci.* 9:12. doi: 10.3389/fnsys.2015.00012

Smith, D. W., Turner, D. A., and Henson, M. M. (2000). Psychophysical correlates of contralateral efferent suppression. I. The role of the medial olivocochlear system in "central masking" in nonhuman primates. *J. Acoust. Soc. Am.* 107, 933–941. doi: 10.1121/1.428274

Srinivasan, S., Keil, A., Stratis, K., Osborne, A. F., Cerwonka, C., Wong, J., et al. (2014). Interaural attention modulated outer hair cell function. *Eur. J. Neurosci.* 40, 3785–3792. doi: 10.1111/ejn.12746

Srinivasan, S., Keil, A., Stratis, K., Woodruff-Carr, K. L., and Smith, D. W. (2012). Effects of cross-modal selective attention on the sensory periphery: cochlear sensitivity is altered by selective attention. *Neuroscience* 223, 325–332. doi: 10.1016/j.neuroscience.2012.07.062

Stinson, M. R., Shaw, E. A. G., and Lawton, B. W. (1982). Estimation of acoustical energy reflectance at the eardrum from measurements of pressure distribution in the human ear canal. *J. Acoust. So.c Am.* 72, 766–773. doi: 10.1121/1.388257

Stuart, A., and Butler, A. K. (2012). Contralateral suppression of transient otoacoustic emissions and sentence recognition in noise in young adults. *J. Am. Acad. Audiol.* 23, 686–696. doi: 10.3766/jaaa.23.9.3

Stuart, A., and Cobb, K. M. (2015). Reliability of measures of transient evoked otoacoustic emissions with contralateral suppression. *J. Comm. Dis.* 58, 35–42. doi: 10.1016/j.jcomdis.2015.09.003

Talmadge, C. L., Long, G. R., Tubis, A., and Dhar, S. (1999). Experimental confirmation of the two-source interference model for the fine structure of distortion product otoacoustic emissions. *J. Acoust. Soc. Am.* 105, 275–292. doi: 10.1121/1.424584

Van Hoesel, R. J., and Clark, G. M. (1997). Psychophysical studies with two binaural cochlear implant subjects. *J. Acoust. Soc. Am.* 102, 495–507. doi: 10.1121/1.419611

Veuillet, E., Duverdy-Bertholon, F., and Collet, L. (1996). Effect of contralateral acoustic stimulation on the growth of click-evoked otoacoustic emissions in humans. *Hear Res.* 93, 128–135. doi: 10.1016/0378-5955(95)00212-x

Wagner, W., Frey, K., Heppelmann, G., Plontke, S. K., and Zenner, H. P. (2008). Speech-in-noise intelligibility does not correlate with efferent olivocochlear reflex in humans with normal hearing. *Acta Otolaryngol.* 128, 53–60. doi: 10.1080/00016480701361954

Wagner, W., and Heyd, A. (2011). Measurement of medial olivocochlear efferent activity in humans: comparison of different distortion product otoacoustic emission-based paradigms. *Otol. Neurotol.* 32, 1379–1388. doi: 10.1097/mao.0b013e31822f1548

Warr, W. B., and Guinan, J. J. (1979). Efferent innervation of the organ of corti: two separate systems. *Brain Res.* 173, 152–155. doi: 10.1016/0006-8993(79)91104-1

Warren, E. H., and Liberman, M. C. (1989). Effects of contralateral sound on auditory-nerve responses. I. Contributions of cochlear efferents. *Hear Res.* 37, 89–104. doi: 10.1016/0378-5955(89)90032-4

Wicher, A., and Moore, B. C. J. (2014). Effect of broadband and narrowband contralateral noise on psychophysical tuning curves and otoacoustic emission. *J. Acoust. Soc. Am.* 135, 2931–2941. doi: 10.1121/1.4871358

Winslow, R. L., and Sachs, M. B. (1988). Single-tone intensity discrimination based on auditor-nerve rate responses in backgrounds of quiet, noise, and with stimulation of the crossed olivocochlear bundle. *Hear Res.* 35, 165–190. doi: 10.1016/0378-5955(88)90116-5

Wittekindt, A., Kaiser, J., and Abel, C. (2014). Attentional modulation of the inner ear: A combined otoacoustic emission and EEG study. *J. Neurosci.* 34, 9995–10002. doi: 10.1523/jneurosci.4861-13.2014

Yasin, I., Drga, V., and Plack, C. J. (2014). Effect of human auditory efferent feedback on cochlear gain and compression. *J. Neurosci.* 34, 15319–15326. doi: 10.1523/jneurosci.1043-14.2014

Zhang, F., Boettcher, F. A., and Sun, X. M. (2007). Contralateral suppression of distortion product otoacoustic emissions: effect of the primary frequency in Dpgrams. *Int. J. Audiol.* 46, 187–195. doi: 10.1080/14992020601164162

Zhao, W., and Dhar, S. (2010). The effect of contralateral acoustic stimulation on spontaneous otoacoustic emissions. *J. Assoc. Res. Otolaryngol.* 11, 53–67. doi: 10.1007/s10162-009-0189-4

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Marrufo-Pérez, Johannesen and Lopez-Poveda. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Neuroscience

BRIEF RESEARCH REPORT
published: 15 April 2021
doi: 10.3389/fnins.2021.654479

Characterization of Developmental Changes in Spontaneous Electrical Activity of Medial Superior Olivary Neurons Before Hearing Onset With a Combination of Injectable and Volatile Anesthesia

Mariano Nicolás Di Guilmi[1]* *and Adrián Rodríguez-Contreras*[2]*

[1] Instituto de Investigaciones en Ingeniería Genética y Biología Molecular, Dr. Héctor N. Torres, INGEBI-CONICET, Buenos Aires, Argentina, [2] Department of Biology, Center for Discovery and Innovation, City College, Institute for Ultrafast Spectroscopy and Lasers, City University of New York, New York, NY, United States

OPEN ACCESS

Edited by:
David Pérez-González,
University of Salamanca, Spain

Reviewed by:
Achim Klug,
University of Colorado, United States
Lars Kunz,
Ludwig Maximilian University of Munich, Germany

***Correspondence:**
Mariano Nicolás Di Guilmi
mndiguilmi@dna.uba.ar
Adrián Rodríguez-Contreras
arodriguezcontreras@ccny.cuny.edu

Specialty section:
This article was submitted to Auditory Cognitive Neuroscience, a section of the journal Frontiers in Neuroscience

Received: 16 January 2021
Accepted: 25 March 2021
Published: 15 April 2021

Citation:
Di Guilmi MN and Rodríguez-Contreras A (2021) Characterization of Developmental Changes in Spontaneous Electrical Activity of Medial Superior Olivary Neurons Before Hearing Onset With a Combination of Injectable and Volatile Anesthesia. Front. Neurosci. 15:654479. doi: 10.3389/fnins.2021.654479

In this work the impact of two widely used anesthetics on the electrical activity of auditory brainstem neurons was studied during postnatal development. Spontaneous electrical activity in neonate rats of either sex was analyzed through a ventral craniotomy in mechanically ventilated pups to carry out patch clamp and multi-electrode electrophysiology recordings in the medial region of the superior olivary complex (SOC) between birth (postnatal day 0, P0) and P12. Recordings were obtained in pups anesthetized with the injectable mix of ketamine/xylazine (K/X mix), with the volatile anesthetic isoflurane (ISO), or in pups anesthetized with K/X mix that were also exposed to ISO. The results of patch clamp recordings demonstrate for the first time that olivary and periolivary neurons in the medial region of the SOC fire bursts of action potentials. The results of multielectrode recordings suggest that the firing pattern of single units recorded in K/X mix is similar to that recorded in ISO anesthetized rat pups. Taken together, the results of this study provide a framework to use injectable and volatile anesthetics for future studies to obtain functional information on the activity of medial superior olivary neurons *in vivo*.

Keywords: ketamine, isoflurane, MNTB, *in vivo* electrophysiology, multi-unit activity

INTRODUCTION

The accurate organization of neuronal circuits is established during development through activity-dependent and activity-independent processes that involve the reorganization and fine-tuning of immature synaptic and cellular networks (Goodman and Shatz, 1993; Hanson and Landmesser, 2004; Kirkby et al., 2013). Before the onset of sensation, spontaneously active cells and signaling mechanisms in sensory organs have been identified as drivers of bursts of neuronal activity that are implicated in activity-dependent refinement of auditory and visual systems in different

vertebrate species (Maffei and Galli-Resta, 1990; Kotak and Sanes, 1995; Lippe, 1995; Jones et al., 2007; Tritsch et al., 2007; Sonntag et al., 2009; Seabrook et al., 2017; Gribizis et al., 2019). In the auditory system of altricial rodents, calcium action potentials (APs) in cochlear inner hair cells (IHCs) initiate minibursts of APs in auditory neurons before the onset of hearing (Tritsch et al., 2007, 2010; Wang et al., 2015). During the prehearing period, IHCs are transiently innervated by direct axo-somatic efferent synaptic contacts from medial olivocochlear (MOC) neurons located in the brainstem superior olivary complex (SOC; Warr and Guinan, 1979; Simmons et al., 1996). MOC innervation is cholinergic (Glowatzki and Fuchs, 2000; Katz et al., 2004; Gómez-Casati et al., 2005) and mediated by postsynaptic acetylcholine nicotinic receptors (nAChR) containing α9 and α10 subunits (Elgoyhen et al., 1994, 2001; Weisstaub et al., 2002; Lipovsek et al., 2012). The nAChR is coupled to the activation of small-conductance calcium-activated SK2 potassium channels expressed in the IHCs, which mediate the hyperpolarization of IHC membrane potential in response to MOC efferent activation (Glowatzki and Fuchs, 2000). Thus, it has been proposed that MOC efferent-mediated inhibition might contribute to pattern trains of IHC calcium APs during the critical developmental period preceding hearing onset (Kros et al., 1998; Glowatzki and Fuchs, 2000; Marcotti et al., 2003; Johnson et al., 2011; Sendin et al., 2014; Moglie et al., 2018).

In recent years, different lines of genetically manipulated mouse models were used to study how modulation of the cochlear pacemaker affects the maturation of central auditory neurons and synapses. In the brainstem medial nucleus of the trapezoid body (MNTB), electrophysiological techniques were employed in slices and *in vivo* recordings of anesthetized α9 nAChR mutant mouse pups (Clause et al., 2014; Di Guilmi et al., 2019). Despite this progress, we identified a major discrepancy in published studies based on the use of different types of anesthetics in different rodent species. The two most widely used anesthetics in brainstem studies are the injectable mix of ketamine/xylazine (K/X mix) in mice (Sonntag et al., 2009; Clause et al., 2014; Di Guilmi et al., 2019), and the volatile anesthetic isoflurane (ISO) in rats (Tritsch et al., 2010; Crins et al., 2011; Sierksma and Borst, 2017). Although different factors are taken into account to choose the type of anesthetic, few studies have addressed its potential effects on *in vivo* spontaneous neuronal activity. For example, although previous studies have identified burst firing units in mice and rats (Sonntag et al., 2009; Tritsch et al., 2010; Clause et al., 2014), only one study in ISO-anesthetized rats reported regular firing units (Tritsch et al., 2010). This raises the possibility that the firing pattern of MNTB and other auditory brainstem neurons could be affected by the type of anesthetic used. Recent works in awake animals do not settle this issue, because optical calcium reporter fluorescence techniques were employed in auditory midbrain and cortex (Babola et al., 2018, 2020; Wang H.-Y. et al., 2020), and when electrophysiology experiments were used, they were performed in the somatosensory cortex of awake mice (Wang Y. et al., 2020).

In this work, we studied neuronal activity with electrophysiology techniques for single unit recording and ensemble multi-unit recording in the SOC of neonate rats. We used K/X mix to induce and maintain rat pups in a stable anesthetized state, while the percent ISO was titrated with precision to study its inhibitory effects on single unit and multiunit activity of medial superior olivary neurons. We also used the two anesthetics independently to monitor the overall activity of a population of neurons in animals at different postnatal ages, and ISO to identify the location of neurons with different firing properties. We found that burst and regular firing units are present in K/X mix anesthetized rat pups, that when ISO is added to K/X mix anesthetized pups it has inhibitory effects on the two types of units, and that some regular firing cells seem to be resistant to the inhibitory effects of maximal doses of ISO. Our results also demonstrate that despite the suppressive effects of anesthetics, the ensemble electrical activity of superior olivary neurons recorded in K/X mix or in ISO-anesthetized rat pups ramps up during development.

MATERIALS AND METHODS

Animal Housing and Breeding

The Institutional Animal Care and Use Committee of the City College of New York specifically reviewed and approved this study. The cohorts of adult Wistar rats used in this study were obtained from a commercial supplier at postnatal age 65 (P65, Charles River). Breeding trios of one male and two females were set for 5 days. At the completion of the breeding period, males were removed from the study while females were housed in pairs for 14 more days, and then individually until they gave birth. Pups were used from the day of birth (P0) to 12 days after birth (P12). Experiments were designed to minimize the number of animals used.

Ventral Surgery and Brain Processing

A total of 85 pups from 37 litters were prepared for surgery to expose the ventral skull for electrophysiology experiments. The surgical procedures used are those described by Rodríguez-Contreras et al. (2008), with minor modifications. Neonate rat pups were initially anesthetized by an intraperitoneal injection of a mixture of ketamine hydrochloride (0.1 mg/g body weight; Ketathesia) and xylazine hydro-chloride (0.005 mg/g body weight; Anased), here referred as K/X mix. After anesthesia induction, animals were tracheotomized, intubated, and mechanically ventilated using a MiniVent type 845 ventilator (Harvard Apparatus). During surgery, anesthesia was carefully monitored on the basis of pedal reflexes, and maintained by supplementary injections of one-third of the initial dose of K/X mix as necessary (Sonntag et al., 2009). For the subset of experiments described in **Figure 2**, animals were initially anesthetized inside an induction chamber with 3% isoflurane (ISO) carried in oxygen, and maintained anesthetized with 1.5% ISO carried in oxygen, delivered via a nose cone or a direct connection to the mechanical ventilator. A small ventral craniotomy (1.5 × 1.5 mm) was performed on each pup, and the brain vascular landscape constituted by the basilar artery and the anterior inferior cerebellar artery was exposed. The dura was carefully removed prior to recording. Body temperature

was controlled during surgery and recording with a heating pad set to 37° (FHC, Maine, United States). After successful recordings were obtained, animals received an overdose injection of Euthasol, and were removed from the setup to be perfused with a solution of 4% paraformaldehyde in 0.1 M phosphate buffer. Brains were removed from the skull and processed for histology to be sectioned in 60 μm thick slices. Brainstem slices were mounted onto glass slides and counterstained with DAPI prior to applying mounting medium and a glass coverslip. Brain sections were imaged with an LSM800 confocal microscope located at the CCNY imaging core facility.

Patch Clamp Recordings

Recordings were made at a depth of 200–600 μm from the pial surface using borosilicate glass pipettes (5–8 MΩ) filled with artificial cerebrospinal fluid (ACSF) containing (in mM): 125 NaCl, 5 KCl, 2 $MgSO_4$, 2 $CaCl_2$, 10 D-glucose, 10 HEPES, adjusted to pH 7.4 with NaOH, or intracellular solution containing (in mM): 126 K-gluconate, 20 KCl, 10 Na_2- phosphocreatine, 4 MgATP, 0.3 Na_2GTP, 0.5 EGTA, and 10 HEPES (310 mmol.kg^{-1}), adjusted to pH 7.2 with KOH. Extracellular potentials were detected in the loose-patch configuration, while intracellular potentials were recorded in the whole-cell configuration. Pipettes were advanced from the brain surface in large steps of 50 μm to a depth of 200 μm using high positive pressure (300 mbar). When in cell search mode, pressure was reduced to 30 mbar and the pipette was moved in 1.5 or 2 μm steps. The pipette tip resistance was monitored with a 10 mV test pulse in voltage clamp or when a shift in the DC was noted in current clamp. When the tip resistance increased suddenly, suggesting that a putative cell was encountered, pressure was released and slight negative pressure of 30–50 mbar was applied until a low resistance seal was formed (10–50 MΩ) for loose-patch recordings, or until a high resistance seal was formed (1–10 GΩ) for whole-cell recordings. High-resistance seals were broken by applying fast bouts of negative pressure. Whole-cell recordings were not corrected for the liquid junction potential. Voltage was recorded with pClamp9 software using an Axopatch 200B amplifier, filtered at 2–5 kHz and sampled at 20 kHz with a Digidata 1440B (MDS Analytical Technologies). Inclusion of Alexa Fluor 594 (1%, Invitrogen) in the pipette solution allowed *post hoc* confirmation of recording sites or recorded cells in paraformaldehyde-fixed brainstem sections.

Multi-Electrode Recordings

Prior to recording silicon probes (polytrode 4 × 4 16-channel arrays, NeuroNexus, A4 × 4–3 mm-50–125-177-A16; ∼1 MΩ) were coated with the lipophilic dye DiI (Invitrogen) to confirm proper targeting to the MNTB and medial region of the SOC in *post hoc* histological analysis. Recordings were made by placing probes with the medial-most shank at coordinates of 300–600 μm rostral and 300 μm lateral to the branch point of the anterior cerebellar artery from the basilar artery, and at a depth of 400–500 μm from the pial surface. All data from silicon probe recordings was acquired in blocks of 10 min, sampled at 24.4 kHz, amplified and digitized in a single-head stage (TDT system III hardware, Tucker-Davis Technologies, Alachua, FL).

Effect of Isoflurane on K/X Mix Anesthetized Pups

ISO was delivered in medical grade oxygen gas via a custom-made tube that connected a vaporizer to the animal ventilator at a rate of 1 L min^{-1}. For single-unit recordings three protocols of ISO delivery were used. The first protocol consisted of acquiring 5 blocks of 10 min each to obtain a baseline, two consecutive blocks in 1.5% isoflurane, and two washout blocks. The second was a staircase protocol that consisted of a single 10 min long recording where APs were quantified in six 100 s bins: baseline (0% ISO), 1.5% ISO, 3.0% ISO, 4.5% ISO and two washout bins (0% ISO). A third protocol was used for units that showed resistance to the staircase protocol and consisted of six 100 s bins: baseline, three bins at 5% ISO, and two washout bins. The effect of ISO was evaluated in multi-electrode recordings by using a procedure similar to the first protocol described above, except that between the baseline recording and the first two blocks in 1.5% ISO, a 90 s period in 1.5% ISO was added.

Data and Statistical Analyses

Unless indicated, data is presented as mean ± sem. In all cases, n indicates the number of units or cells tested, with the exception of **Figure 3** where n indicates the number of animals. Extracellular single-unit data was filtered in two steps. First, to eliminate 60 cycle electrical interference (0.6 Hz, 3 dB bandwidth); and second, with a high-pass Bessel (8-pole) filter with a −3 dB cutoff of 5 Hz using Clampfit 10.6 (Molecular Devices). All data was analyzed using NeuroMatic in Igor Pro software (WaveMetrics; Rothman and Silver, 2018). APs were detected using a positive threshold set to six times the standard deviation of the baseline noise. Single-unit recordings where an obvious gradual decrease or rundown in spike amplitude occurred were excluded from analysis. Multielectrode multi-unit data was exported as NEX files for analysis in Plexon software (Offline sorter 4.4 and Neuroexplorer). APs were detected using a negative threshold set to six times the standard deviation of the baseline noise. Multiunit activity (MUA) collected in every channel was considered. All statistical tests were performed with Statistica 7.0 software (Stat Soft; RRID:SCR_014213). Non-parametric Kruskall-Wallis (K-W) test was used. Values of $p < 0.05$ were considered significant.

Drugs and Reagents

All drugs and reagents were purchased from Sigma-Aldrich (RRID:SCR_008988).

RESULTS

Isoflurane Inhibits Burst and Regular Firing Units in the Ventral Brainstem of K/X Mix-Anesthetized Rat Pups

Single-unit recordings were performed in five K/X mix-anesthetized rat pups before hearing onset, and the effects of 1.5% isoflurane were evaluated on the firing properties of six single-units (one unit from a P3 pup and five units from four P6 pups). Two of the units showed burst firing

and four units showed regular firing patterns under baseline conditions (exemplar recordings are shown in **Figures 1A–C**). These recorded units had variable responses to ISO. The activity of the two burst firing units was inhibited to 96 and 99.6% compared to baseline and did not recover during washout. The activity of three out of the four regular firing units was partially inhibited to 34 ± 21% compared to baseline, and only in one unit recovered to baseline levels during washout. Lastly, one of the four regular firing units slightly increased its activity by 2% in ISO compared to baseline, but its activity was suppressed to 19% during washout. These inhibitory and potentiating effects of ISO on single-unit firing properties were observed in the inter-spike-interval histograms shown in **Figures 1D–F**. Small changes in the waveform of recorded units were also noted (insets in **Figures 1D–F**).

To address the variation of 1.5% ISO effects on recorded units, additional experiments were performed on thirteen single-units recorded from two P3 pups, two P6 pups, and three P9 pups by increasing the ISO dose from 0 to 4.5% in 1.5% steps. It was found that two P3 units, two P6 units and four P9 units showed clear signs of dose-dependent inhibition at 4.5% ISO compared to baseline. These cells did not recover during the washout period (**Figure 1G**). In a total of five additional single-units from pups of all ages, evidence of potentiation or resistance to 4.5% ISO was found. In three of these units, a small inhibitory effect was observed during the washout period (**Figure 1H**). Three of these

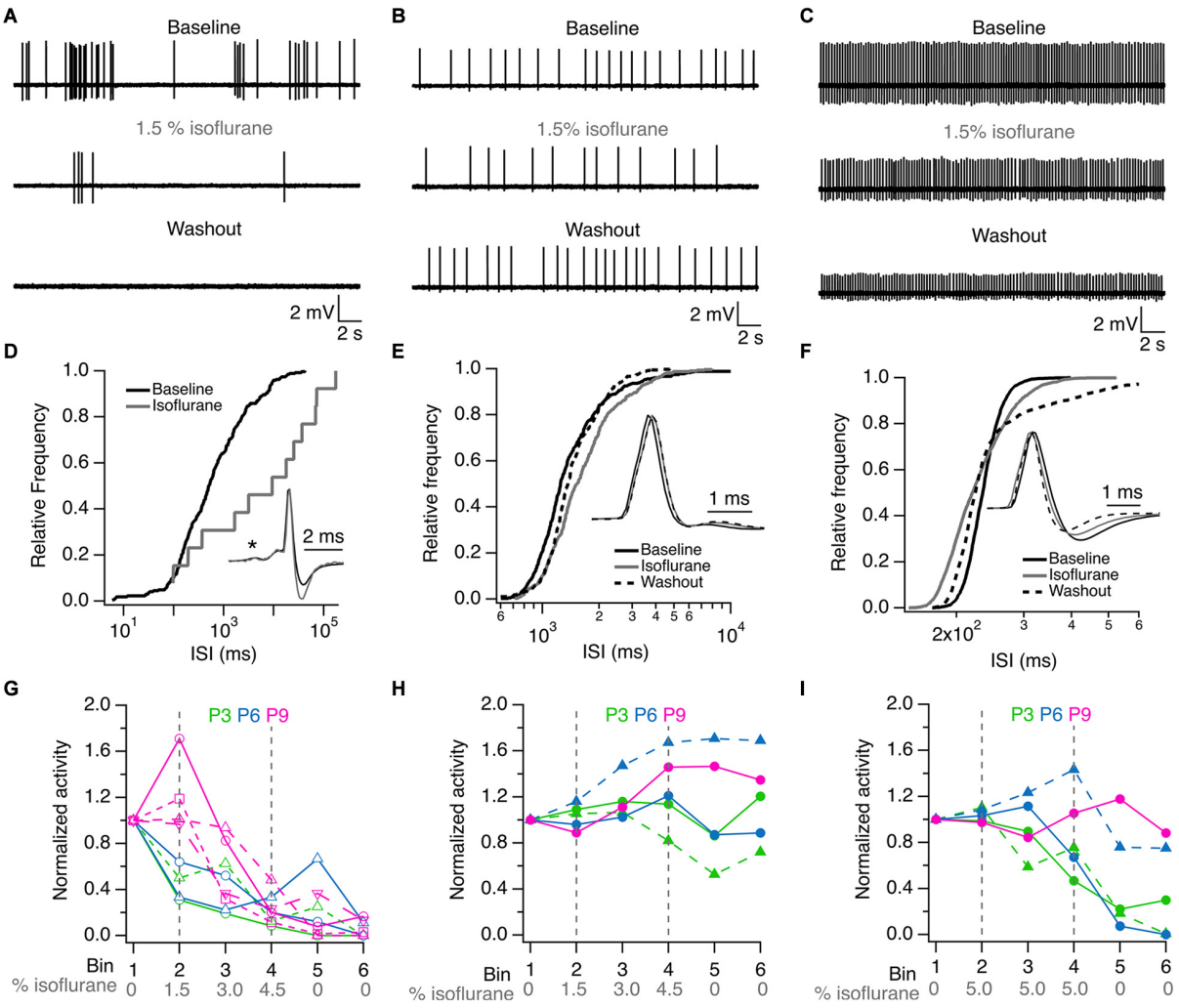

FIGURE 1 | ISO inhibit single unit activity in K/X mix anesthetized pups. **(A)** Burst firing unit from a P6 rat. **(B)** Regular firing unit from a P3 rat. **(C)** Regular firing unit from a P6 rat. **(D)** Effect of ISO on the inter-spike interval (ISI) distribution of single unit shown in **(A)**. **(E)** Effect of ISO on ISI distribution of single unit shown in **(B)**. **(F)** Effect of isoflurane on ISI distribution of single unit shown in **(C)**. **(G)** Inhibitory effect of ISO on single unit activity recorded at P3 (green), P6 (blue) and P9 (magenta). **(H)** ISO resistant single units recorded at P3, P6 and P9. **(I)** Exposure to 5% ISO inhibited 3 of 5 single units shown in **(H)**. ISI = inter-spike interval. Asterisk in D indicates the pre-spike in this complex waveform. Inset waveforms in **(D)** are averages of 269 and 11 action potentials in baseline and isoflurane, respectively. Inset waveforms in **(E)** are averages of 363, 330, and 392 action potentials in baseline, isoflurane and washout, respectively. Inset waveforms in **(F)** are averages of 2,538, 2,599, and 2,046 action potentials in baseline, isoflurane and washout, respectively. Lines and symbols in **(H,I)** identify the same single units under different isoflurane conditions. Vertical dashed lines in **(G–I)** indicate the onset and offset of isoflurane application.

units were inhibited by prolonged exposure to 5% ISO and did not recover during the washout period. The remaining two units from P6 and P9 pups showed signs of resistance and potentiation even at such high dosage (**Figure 1I**).

Altogether, these experiments show that burst and regular firing cells are present in K/X mix-anesthetized rat pups, that most burst and regular firing units are inhibited permanently by ISO, while some regular firing units appear to be resistant to inhibition and show indications of potentiation after ISO is delivered to K/X mix-anesthetized rat pups. Resistance and potentiation to ISO was observed more frequently in P6 and P9 rat pups.

Heterogeneous Population of Single-Units: Inactive Versus Active Neurons

It was noted that from a total of nineteen recorded single-units, five units fired in a burst type pattern, while the rest fired in a regular spike pattern. Furthermore, three of the burst firing units did not show complex waveforms adjudicated to MNTB neurons. These observations raised the possibility that other SOC neurons may fire burst type or regular type patterns. One caveat of loose-seal patch clamp recordings is that cells cannot be identified, unless iontophoretic methods are used to label them (Pinault, 1996; Cid and de la Prida, 2019). As an alternative to identify the morphology and location of recorded cells, whole-cell current clamp recordings were performed in forty ISO anesthetized pups between ages P0 and P11 (**Figure 2**). Active and inactive cells were identified in current clamp recordings. Active cells fired APs in burst ($n = 10$) or regular type ($n = 5$) patterns, while inactive cells did not fire APs at rest ($n = 25$; **Figure 2A**). Two kinds of inactive cells were further identified based on their response to depolarizing current injection. Some cells fired APs ($n = 13$), while other cells did not ($n = 12$; **Figure 2B**). Therefore, inactive cells were classified as able or unable to fire APs (**Figure 2C**). After histological processing, the morphology and location of twenty-three recorded cells was confirmed in the medial region of the SOC (**Figure 2D**). Eight burst firing cells were observed within and outside the boundaries of SOC nuclei such as the MNTB and the MSO/SPN; two regular spiking cells were observed in the dorsal region outside the boundary of the MNTB; two inactive cells were observed within the boundaries of the MNTB; and eleven inactive cells were found at ventral, medial, lateral and dorsal locations outside the boundaries of the MNTB and MSO/SPN.

In sum, these experiments show that ISO-anesthetized rat pups have active and inactive neurons in the SOC. Several cells located in the medial region of the SOC fire APs in burst or regular patterns. However, while burst-firing cells were observed within and outside the boundaries of SOC nuclei, regular firing cells were observed only outside the boundaries of SOC nuclei. Some inactive cells are capable of firing action potentials upon depolarization of their membrane potential, while others cannot fire action potentials suggesting irreversible inhibition or a different cellular phenotype (i.e., non-neuronal glial).

Multiunit Activity Increases During Postnatal Development

Next, the effect of the injectable and volatile anesthetics on the ensemble activity of the medial SOC was evaluated separately

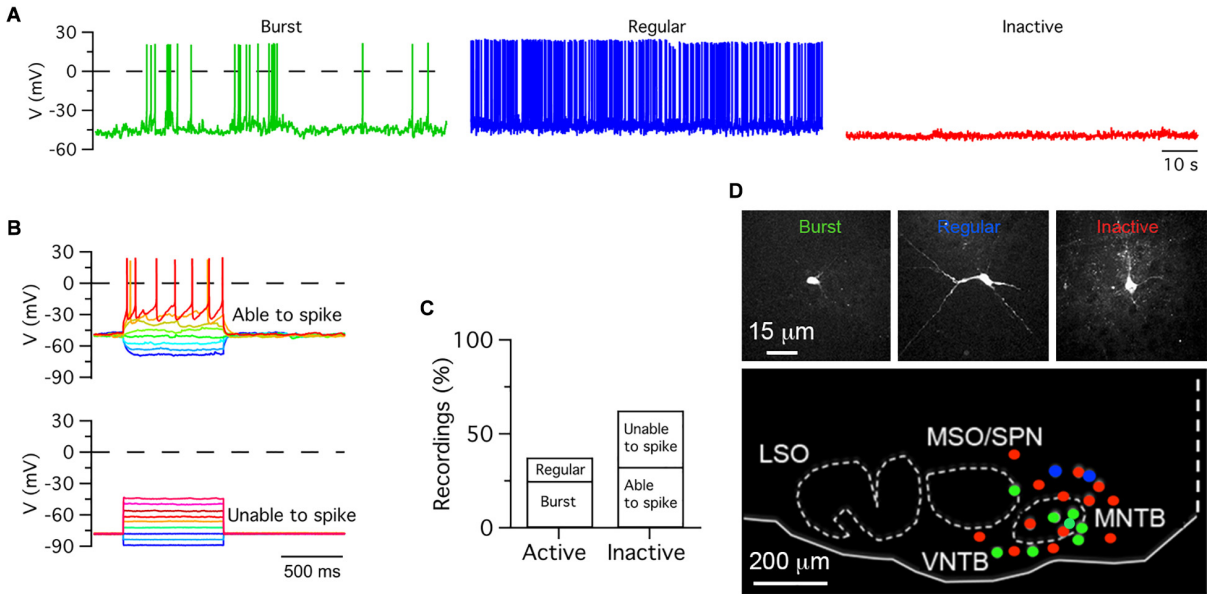

FIGURE 2 | Heterogeneous electrical activity in medial olivocochlear neurons. **(A)** Exemplar patch-clamp whole cell recordings of burst firing, regular firing and inactive cells obtained from pups anesthetized with 1.5% ISO. **(B)** Inactive cells were subjected to hyperpolarizing and depolarizing voltage steps. Depolarization steps triggered action potentials in some cells but not in others. **(C)** Percent of active and inactive cells. **(D)** Exemplar images of cells shown in **(A)** (top row). Distribution of recorded cells in the auditory brainstem in neonate pups (bottom). Recording traces were not corrected for the liquid junction potential.

with multi-electrode probes in pups aged between P0 and P12 ($n = 29$; **Figure 3A**). Despite the large variability of spontaneous MUA within each developmental group, the overall MNTB spontaneous activity ramped up during development under K/X mix ($n = 14$; **Figures 3B,C**, left panel) or ISO ($n = 15$; **Figures 3B,C** right panel). It is important to note that the mean firing rate was similar under both conditions. Statistical analysis showed only differences under ISO displaying lower activity at P0 compared to P9 (K-W, $p = 0.04$; multiple comparisons, $p < 0.05$; $n = 3$). Taken together, these results suggest that the overall level of spontaneous electrical activity increases during postnatal development and that the maximal ensemble firing rate was similar between the two anesthetics.

Developmental Profile of ISO Inhibition

To examine whether ISO may act over the anesthetic effect of K/X mix and considering that both anesthetics have different synaptic targets (MacDonald et al., 1987; Wang Y. et al., 2020), we implemented the recording protocol illustrated in **Figure 4A**. **Figure 4B** displays three examples of MUA at different developmental ages (P3, P6, and P9) along the experimental protocol. In all cases, application of 1.5% ISO depressed the spiking activity, and a partial or total washout could be observed. The activity level was quantified after application of 1.5% ISO relative to K/X mix baseline (**Figure 4C**). P0 was excluded from the statistical analysis due to the activity being too small and the estimation of inhibition may not be reliable. Statistical analysis showed that ISO at 10 min had a greater initial inhibition at P6 in comparison with P12 (K-W, $p = 0.007$; multiple comparisons, $p < 0.05$; $n = 3$). In conjunction with loose-seal patch clamp recordings (**Figure 1**), these experiments show that the inhibitory effect of ISO is heterogeneous before hearing onset.

DISCUSSION

General anesthetics are widely used for surgical procedures in animal models and sometimes are required during *in vivo* experiments. For example, recent studies have targeted dorsally accessible brain structures like the inferior colliculus (IC) and auditory cortex (AC) in awake transgenic mice that had been exposed to anesthetics during preparatory surgery

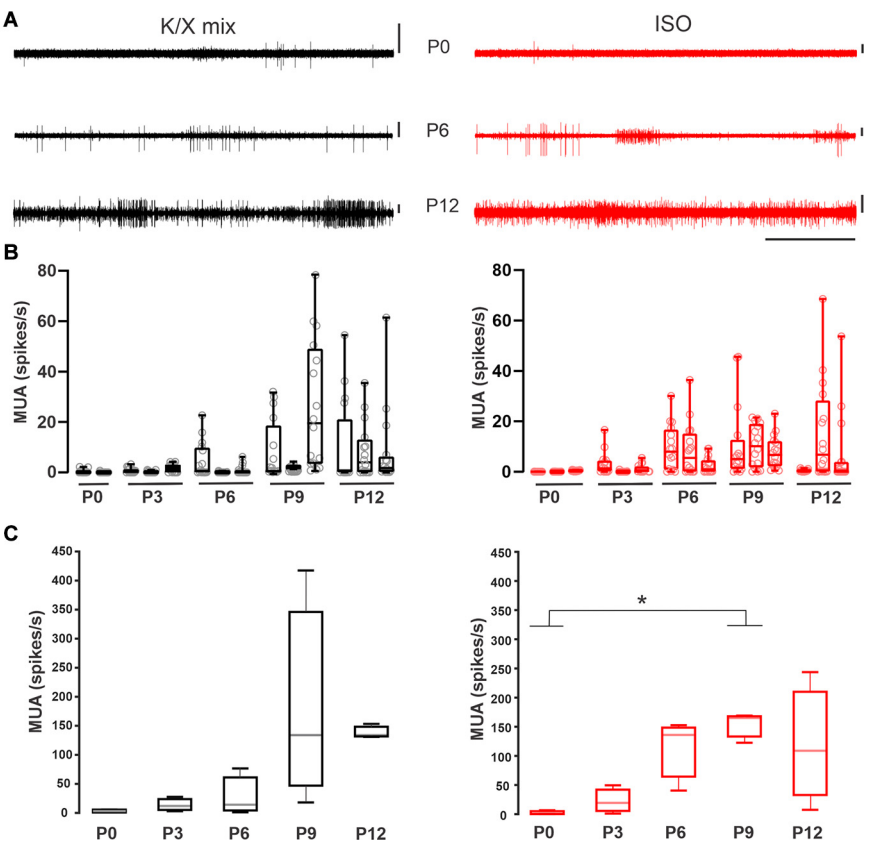

FIGURE 3 | Increase in multiunit activity in the MNTB before hearing onset. **(A)** Representative recordings acquired with a polytrode in the medial portion of the SOC of neonate rats. Exemplar recording at P0 (upper), P6 (middle), and P12 (bottom panel) under K/X mix (left) or ISO (right panel). Vertical scale bars = 100 μV, horizontal scale bar = 5 s. Note the increase in the activity level along development. **(B)** Box plots for different animals illustrating the MUA frequency in each electrode (empty circles). **(C)** Quantification of the averaged multiunit activity per developmental stage. Boxes represent interquartile range between 25 and 75%. Whiskers indicate the minimum and maximum of all data. Inside line represents the median. Statistical analysis show differences under ISO displaying lower activity at P0 compared to P9 (K-W, $p = 0.04$; multiple comparisons, * indicates multiple comparisons, $p < 0.05$; $n = 3$).

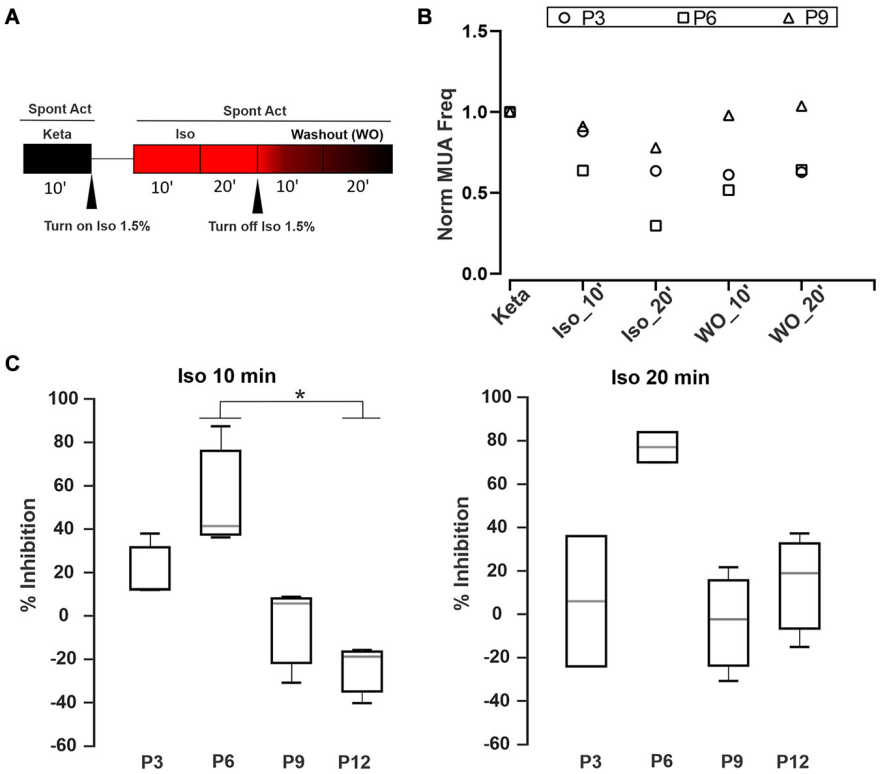

FIGURE 4 | Time-dependent inhibition of Isoflurane. **(A)** Schematic diagram of the pharmacological protocol for testing Isoflurane inhibition. The first step was to record the spontaneous activity during 10 min under K/X mix (black box). Then we applied Isoflurane 1.5% for 90 s followed by two recording rounds of 10 min to obtain MUA after 10 or 20 min, respectively (red boxes). After that, ISO was turned off to evaluate the washout for the next 20 min. **(B)** Three examples showing the ISO inhibition and its washout at different developmental ages (P3, P6, and P9). Normalized MUA frequency relative to the K/X mix state (Norm MUA freq). **(C)** Percentage of Inhibition (MUA frequency under ISO/MUA frequency under K/X mix) after 10 (left) or 20 min (right) with ISO at different developmental stages. Statistical analysis show that ISO at 10 min had a greater inhibition at P6 in comparison with P12 (K-W, $p = 0.007$; multiple comparisons, * indicates multiple comparisons, $p < 0.05$; $n = 3$).

(Babola et al., 2018, 2020). However, if ventrally accessible brainstem structures are targeted to perform imaging and/or electrophysiological recordings, a general anesthetic must be used to perform recordings in animals that are mechanically ventilated (Rodríguez-Contreras et al., 2014; Di Guilmi et al., 2019), or that are immobilized to facilitate accurate electrode targeting (Sonntag et al., 2009; Clause et al., 2014). In this work we studied the impact of K/X mix and ISO on the spontaneous electrical activity of medial superior olivary neurons in neonate rodents.

This study provides three main findings that are relevant to understand the role of medial superior olivary neurons in auditory system development. First, to the best of our knowledge, this is the first study that compares K/X mix and ISO in the same rodent species. The firing pattern of single units recorded in K/X mix was similar to that reported in ISO anesthetized rat pups (**Figure 1**; Tritsch et al., 2010). Furthermore, this is the first study that identified burst-firing neurons outside the MNTB, which due to their location were mapped to the medial portion of the SOC (**Figure 2**). Second, it was found that ISO displayed a large scatter on its suppressing effects on electrical activity of auditory brainstem neurons. In pups that were pre-anesthetized with K/X mix, around two thirds of recorded cells were inhibited by 1.5% isoflurane, the isoflurane EC50 estimated under controlled conditions (**Figures 1**, **4**; Wang Y. et al., 2020). However, an estimated one third of recorded units, all of which had a regular firing pattern, were resistant to 1.5% and were inhibited by 5% isoflurane delivered over 20 min. Third, this work showed that electrical ensemble activity of medial superior olivary neurons ramps up during development before hearing onset, with a similar developmental profile under both anesthetics (**Figure 3**).

One caveat of the present study is that we were not able to study the effect of different concentrations of K/X mix on auditory brainstem neuronal activity. Future studies could first induce and maintain an anesthetic state with ISO, while K/X mix is delivered and titrated with precision via intravascular infusion. Another limitation of the present study is that the spatial location of neurons with different firing properties was inferred after intense *post hoc* histological analysis, and the functional properties of those cells could not be studied in the same animal. We consider that our results shed light on the use of anesthetics as tools for future studies that enable patch-clamp electrophysiology combined with imaging studies of medial superior olivary neurons *in vivo*. The approach would be to use K/X mix to induce a stable anesthetic state and ISO to probe

neuron inhibition on burst and regular firing neurons, which are predicted to have different responses to ISO. A combined optical calcium reporter imaging and electrophysiology approach would be desirable and useful to study MOC neurons, whose cell bodies are diffusely located in the ventral nucleus of the trapezoid body (VNTB) (Warr, 1975; Warr and Guinan, 1979; Guinan et al., 1983) and could be labeled with fluorescent reporters using genetic approaches (Cadenas et al., 2020).

Lastly, it is important to consider how the present study may advance our understanding of the mechanisms of anesthetic action throughout life, since neurons in the medial superior olive have also been studied in adult rodents anesthetized with K/X mix to investigate auditory processing (Lorteije et al., 2009; van der Heijden et al., 2013). It is noted that the K/X mix and ISO have different synaptic targets [i.e., postsynaptic NMDA receptors (MacDonald et al., 1987) or presynaptic effect (Wang Y. et al., 2020), respectively]. In agreement with previous reports, synaptic depressant percentage depended on ISO concentration and the developmental age of pups (Wu et al., 2004; Wang Y. et al., 2020). Different reasons related to the plastic changes of auditory synapses (Wu et al., 2004), the timing of hearing onset (Qiu et al., 2020), and the fact that the principal target of ketamine is the NMDA receptor (MacDonald et al., 1987) which declined to very low levels at the end of the second postnatal week (Taschenberger and von Gersdorff, 2000; Joshi and Wang, 2002), may explain the results of this study. Inhaled anesthetics diffuse from the alveoli into arterial and capillary blood, and are assumed to equilibrate rapidly with the well-perfused central nervous system (Hemmings et al., 2005). Thus, the heterogeneity of inhibitory effects observed in the results of this study, may also be explained by the hypothesis that the physical and functional properties of the vasculature, the diffusive properties of the extracellular space between vessels and neurons, and the clearance mechanisms of extracellular fluid available during this developmental window may be important factors that determine the access of anesthetics to target neurons in the brain (Nicholson et al., 2000; Shi and Rodríguez-Contreras, 2016; Mestre et al., 2020). A major difference in the effects of ISO in adult animals compared to neonates is the induction of slow waves of neuronal activity, an effect not described in neonates (Golshani et al., 2009; Brier et al., 2019). In this study, we observed that a group of recorded cells were inactive. One of the responsible factors of neuronal excitability is the resting membrane potential (Vm). Volatile anesthetics induced negative potentials on cortical cells (Petersen et al., 2003) and the hyperpolarization of postsynaptic MNTB neurons in brain slices (Wang Y. et al., 2020). Rusu and Borst (2010) demonstrated in slices containing the MNTB that Vm decreases during the first postnatal week becoming progressively more negative. During the first two postnatal weeks, the heterogeneity in the intrinsic properties maturation is larger than in adulthood, and probably it is a substrate for our observations. However, the effect of ISO on the Vm of these cells is under debate (Wu et al., 2004; Wang Y. et al., 2020) and a further study of molecular mechanisms during development is necessary to clarify the results.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The animal study was reviewed and approved by the Institutional Animal Care and Use Committee of the City College of New York.

AUTHOR CONTRIBUTIONS

AR-C designed and performed the patch-clamp, multielectrode, and histology experiments. MD designed and performed the multi-electrode experiments. MD and AR-C analyzed the data and wrote and edited the manuscript. Both authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the National Institutes of Health Grant SC1DC015907 to AR-C. MD received the Company of Biologists traveling fellowship DEVTF-160703 and a "Bec.Ar" fellowship for a short stay supported by the Argentinian Government.

ACKNOWLEDGMENTS

We would like to acknowledge the Science Division Core Imaging Facility and Dr. Jonathan Levitt for providing access to a microtome to conduct histology procedures.

REFERENCES

Babola, T. A., Kersbergen, C. J., Wang, H. C., and Bergles, D. E. (2020). Purinergic signaling in cochlear supporting cells reduces hair cell excitability by increasing the extracellular space. *Elife* 9, 1–30. doi: 10.7554/eLife.52160

Babola, T. A., Li, S., Gribizis, A., Lee, B. J., Issa, J. B., Wang, H. C., et al. (2018). Homeostatic Control of Spontaneous Activity in the Developing Auditory System. *Neuron* 99, 511–524. doi: 10.1016/j.neuron.2018.07.004

Brier, L. M., Landsness, E. C., Snyder, A. Z., Wright, P. W., Baxter, G. A., Bauer, A. Q., et al. (2019). Separability of calcium slow waves and functional connectivity during wake, sleep, and anesthesia. *Neurophotonics* 6:035002.

Cadenas, L. T., Fischl, M. J., and Weisz, C. J. C. (2020). Synaptic inhibition of medial olivocochlear efferent neurons by neurons of the medial nucleus of the trapezoid body. *J. Neurosci.* 40, 509–525. doi: 10.1523/JNEUROSCI.1288-19.2019

Cid, E., and de la Prida, L. M. (2019). Methods for single-cell recording and labeling in vivo. *J. Neurosci. Methods* 325:108354. doi: 10.1016/j.jneumeth.2019.108354

Clause, A., Kim, G., Sonntag, M., Weisz, C. J. C., Vetter, D. E., Rubsamen, R., et al. (2014). The precise temporal pattern of prehearing spontaneous activity

is necessary for tonotopic map refinement. *Neuron* 82, 822–835. doi: 10.1016/j.neuron.2014.04.001

Crins, T. T., Rusu, S. I., Rodriguez-Contreras, A., and Borst, J. G. (2011). Developmental changes in short-term plasticity at the rat calyx of Held synapse. *J. Neurosci.* 31, 11706–11717. doi: 10.1523/JNEUROSCI.1995-11.2011

Di Guilmi, M. N., Boero, L. E., Castagna, V. C., Rodríguez-Contreras, A., Wedemeyer, C., Gómez-Casati, M. E., et al. (2019). Strengthening of the Efferent Olivocochlear System Leads to Synaptic Dysfunction and Tonotopy Disruption of a Central Auditory Nucleus. *J. Neurosci.* 39, 7037–7048. doi: 10.1523/JNEUROSCI.2536-18.2019

Elgoyhen, A. B., Johnson, D. S., Boulter, J., Vetter, D. E., and Heinemann, S. (1994). Alpha 9: an acetylcholine receptor with novel pharmacological properties expressed in rat cochlear hair cells. *Cell* 79, 705–715. doi: 10.1016/0092-8674(94)90555-x

Elgoyhen, A. B., Vetter, D. E., Katz, E., Rothlin, C. V., Heinemann, S. F., and Boulter, J. (2001). Alpha10: a determinant of nicotinic cholinergic receptor function in mammalian vestibular and cochlear mechanosensory hair cells. *Proc. Natl. Acad. Sci. U. S. A.* 98, 3501–3506. doi: 10.1073/pnas.051622798

Glowatzki, E., and Fuchs, P. A. (2000). Cholinergic synaptic inhibition of inner hair cells in the neonatal mammalian cochlea. *Science* 288, 2366–2368. doi: 10.1126/science.288.5475.2366

Golshani, P., Goncalves, J. T., Khoshkhoo, S., Mostany, R., Smirnakis, S., and Portera-Cailliau, C. (2009). Internally mediated developmental desynchronization of neocortical network activity. *J. Neurosci.* 29, 10890–10899. doi: 10.1523/jneurosci.2012-09.2009

Gómez-Casati, M. E., Fuchs, P. A., Elgoyhen, A. B., and Katz, E. (2005). Biophysical and pharmacological characterization of nicotinic cholinergic receptors in rat cochlear inner hair cells. *J. Physiol.* 566, 103–118. doi: 10.1113/jphysiol.2005.087155

Goodman, C. S., and Shatz, C. J. (1993). Developmental mechanisms that generate precise patterns of neuronal connectivity. *Cell* 72, 77–98. doi: 10.1016/s0092-8674(05)80030-3

Gribizis, A., Ge, X., Daigle, T. L., Ackman, J. B., Zeng, H., Lee, D., et al. (2019). Visual Cortex Gains Independence from Peripheral Drive before Eye Opening. *Neuron* 104, 711–723. doi: 10.1016/j.neuron.2019.08.015

Guinan, J. J., Warr, W. B., and Norris, B. E. (1983). Differential olivocochlear projections from lateral versus medial zones of the superior olivary complex. *J. Comp. Neurol.* 221, 358–370. doi: 10.1002/cne.902210310

Hanson, M. G., and Landmesser, L. T. (2004). Normal patterns of spontaneous activity are required for correct motor axon guidance and the expression of specific guidance molecules. *Neuron* 43, 687–701. doi: 10.1016/j.neuron.2004.08.018

Hemmings, H. C., Akabas, M. H., Goldstein, P. A., Trudell, J. R., Orser, B. A., and Harrison, N. L. (2005). Emerging molecular mechanisms of general anesthetic action. *Trends Pharmacol. Sci.* 26, 503–510. doi: 10.1016/j.tips.2005.08.006

Johnson, S. L., Eckrich, T., Kuhn, S., Zampini, V., Franz, C., Ranatunga, K. M., et al. (2011). Position-dependent patterning of spontaneous action potentials in immature cochlear inner hair cells. *Nat. Neurosci.* 14, 711–717. doi: 10.1038/nn.2803

Jones, T. A., Leake, P. A., Snyder, R. L., Stakhovskaya, O., and Bonham, B. (2007). Spontaneous discharge patterns in cochlear spiral ganglion cells before the onset of hearing in cats. *J. Neurophysiol.* 98, 1898–1908. doi: 10.1152/jn.00472.2007

Joshi, I., and Wang, L.-Y. (2002). Developmental profiles of glutamate receptors and synaptic transmission at a single synapse in the mouse auditory brainstem. *J. Physiol.* 540, 861–873. doi: 10.1113/jphysiol.2001.013506

Katz, E., Elgoyhen, A. B., Gomez-Casati, M. E., Knipper, M., Vetter, D. E., Fuchs, P. A., et al. (2004). Developmental regulation of nicotinic synapses on cochlear inner hair cells. *J. Neurosci.* 24, 7814–7820. doi: 10.1523/jneurosci.2102-04.2004

Kirkby, L. A., Sack, G. S., Firl, A., and Feller, M. B. (2013). A role for correlated spontaneous activity in the assembly of neural circuits. *Neuron* 80, 1129–1144. doi: 10.1016/j.neuron.2013.10.030

Kotak, V. C., and Sanes, D. H. (1995). Synaptically evoked prolonged depolarizations in the developing auditory system. *J. Neurophysiol.* 74, 1611–1620. doi: 10.1152/jn.1995.74.4.1611

Kros, C. J., Ruppersberg, J. P., and Rusch, A. (1998). Expression of a potassium current in inner hair cells during development of hearing in mice. *Nature* 394, 281–284. doi: 10.1038/28401

Lipovsek, M., Im, G. J., Franchini, L. F., Pisciottano, F., Katz, E., Fuchs, P. A., et al. (2012). Phylogenetic differences in calcium permeability of the auditory hair cell cholinergic nicotinic receptor. *Proc. Natl. Acad. Sci. U. S. A.* 109, 4308–4313. doi: 10.1073/pnas.1115488109

Lippe, W. R. (1995). Relationship between frequency of spontaneous bursting and tonotopic position in the developing avian auditory system. *Brain Res.* 703, 205–213. doi: 10.1016/0006-8993(95)01096-3

Lorteije, J. A. M., Rusu, S. I., Kushmerick, C., and Borst, J. G. G. (2009). Reliability and precision of the mouse calyx of Held synapse. *J. Neurosci.* 29, 13770–13784. doi: 10.1523/jneurosci.3285-09.2009

MacDonald, J. F., Miljkovic, Z., and Pennefather, P. (1987). Use-dependent block of excitatory amino acid currents in cultured neurons by ketamine. *J. Neurophysiol.* 58, 251–266. doi: 10.1152/jn.1987.58.2.251

Maffei, L., and Galli-Resta, L. (1990). Correlation in the discharges of neighboring rat retinal ganglion cells during prenatal life. *Proc. Natl. Acad. Sci. U. S. A.* 87, 2861–2864. doi: 10.1073/pnas.87.7.2861

Marcotti, W., Johnson, S. L., Rusch, A., and Kros, C. J. (2003). Sodium and calcium currents shape action potentials in immature mouse inner hair cells. *J. Physiol.* 552, 743–761. doi: 10.1113/jphysiol.2003.043612

Mestre, H., Mori, Y., and Nedergaard, M. (2020). The Brain's Glymphatic System: current Controversies. *Trends Neurosci.* 43, 458–466. doi: 10.1016/j.tins.2020.04.003

Moglie, M. J., Fuchs, P. A., Elgoyhen, A. B., and Goutman, J. D. (2018). Compartmentalization of antagonistic Ca(2+) signals in developing cochlear hair cells. *Proc. Natl. Acad. Sci. U. S. A.* 115, E2095–E2104.

Nicholson, C., Chen, K. C., Hrabětová, S., and Tao, L. (2000). Diffusion of molecules in brain extracellular space: theory and experiment. *Prog. Brain Resin.* 125, 129–154. doi: 10.1016/S0079-6123(00)25007-3

Petersen, C. C. H., Hahn, T. T. G., Mehta, M., Grinvald, A., and Sakmann, B. (2003). Interaction of sensory responses with spontaneous depolarization in layer 2/3 barrel cortex. *Proc. Natl. Acad. Sci. U. S. A.* 100, 13638–13643. doi: 10.1073/pnas.2235811100

Pinault, D. (1996). A novel single-cell staining procedure performed in vivo under electrophysiological control: morpho-functional features of juxtacellularly labeled thalamic cells and other central neurons with biocytin or Neurobiotin. *J. Neurosci. Methods* 65, 113–136. doi: 10.1016/0165-0270(95)00144-1

Qiu, J., Singh, P., Pan, G., de Paolis, A., Champagne, F. A., Liu, J., et al. (2020). Defining the relationship between maternal care behavior and sensory development in Wistar rats: auditory periphery development, eye opening and brain gene expression. *PLoS One* 15:e0237933. doi: 10.1371/journal.pone.0237933

Rodríguez-Contreras, A., Shi, L., and Fu, B. M. (2014). A Method to Make a Craniotomy on the Ventral Skull of Neonate Rodents. *J. Vis. Exp.* 87, 51350. doi: 10.3791/51350

Rodríguez-Contreras, A., van Hoeve, J. S. S., Habets, R. L. P., Locher, H., and Borst, J. G. G. (2008). Dynamic development of the calyx of Held synapse. *Proc. Natl. Acad. Sci. U. S. A.* 105, 5603–5608. doi: 10.1073/pnas.0801395105

Rothman, J. S., and Silver, R. A. (2018). NeuroMatic: an Integrated Open-Source Software Toolkit for Acquisition, Analysis and Simulation of Electrophysiological Data. *Front. Neuroinform.* 12:14. doi: 10.3389/fninf.2018.00014

Rusu, S. I., and Borst, J. G. (2010). Developmental changes in intrinsic excitability of principal neurons in the rat medial nucleus of the trapezoid body. *Dev. Neurobiol.* 71, 284–295. doi: 10.1002/dneu.20856

Seabrook, T. A., Burbridge, T. J., Crair, M. C., and Huberman, A. D. (2017). Architecture, Function, and Assembly of the Mouse Visual System. *Annu. Rev. Neurosci.* 40, 499–538. doi: 10.1146/annurev-neuro-071714-033842

Sendin, G., Bourien, J., Rassendren, F., Puel, J. L., and Nouvian, R. (2014). Spatiotemporal pattern of action potential firing in developing inner hair cells of the mouse cochlea. *Proc. Natl. Acad. Sci. U. S. A.* 111, 1999–2004. doi: 10.1073/pnas.1319615111

Shi, L., and Rodríguez-Contreras, A. (2016)."In vivo two-photon imaging measuring the blood-brain barrier permeability during early postnatal brain development in rodent," in *Proc. SPIE 9712, Multiphoton Microscopy in the Biomedical Sciences XVI*, eds A. Periasamy, P. T. C. So, and K. König (Washington: SPIE), doi: 10.1117/12.2219571

Sierksma, M. C., and Borst, J. G. G. (2017). Resistance to action potential depression of a rat axon terminal in vivo. *Proc. Natl. Acad. Sci. U. S. A.* 114, 4249–4254. doi: 10.1073/pnas.1619433114

Simmons, D. D., Mansdorf, N. B., and Kim, J. H. (1996). Olivocochlear innervation of inner and outer hair cells during postnatal maturation: evidence for a waiting period. *J. Comp. Neurol.* 370, 551–562. doi: 10.1002/(sici)1096-9861(19960708)370:4<551::aid-cne10>3.0.co;2-m

Sonntag, M., Englitz, B., Kopp-Scheinpflug, C., and Rubsamen, R. (2009). Early postnatal development of spontaneous and acoustically evoked discharge activity of principal cells of the medial nucleus of the trapezoid body: an in vivo study in mice. *J. Neurosci.* 29, 9510–9520. doi: 10.1523/jneurosci.1377-09.2009

Taschenberger, H., and von Gersdorff, H. (2000). Fine-tuning an auditory synapse for speed and fidelity: developmental changes in presynaptic waveform, EPSC kinetics, and synaptic plasticity. *J. Neurosci.* 20, 9162–9173. doi: 10.1523/JNEUROSCI.20-24-09162.2000

Tritsch, N. X., Rodríguez-contreras, A., Crins, T. T. H., Wang, H. C., Borst, J. G., Bergles, D. E., et al. (2010). Calcium action potentials in hair cells pattern auditory neuron activity before hearing onset. *Nat. Neurosci.* 13, 1050–1052. doi: 10.1038/nn.2604

Tritsch, N. X., Yi, E., Gale, J. E., Glowatzki, E., and Bergles, D. E. (2007). The origin of spontaneous activity in the developing auditory system. *Nature* 450, 50–55. doi: 10.1038/nature06233

van der Heijden, M., Lorteije, J. A. M., Plauska, A., Roberts, M. T., Golding, N. L., and Borst, J. G. G. (2013). Directional hearing by linear summation of binaural inputs at the medial superior olive. *Neuron* 78, 936–948. doi: 10.1016/j.neuron.2013.04.028

Wang, H. C., Lin, C. C., Cheung, R., Zhang-Hooks, Y., Agarwal, A., Ellis-Davies, G., et al. (2015). Spontaneous Activity of Cochlear Hair Cells Triggered by Fluid Secretion Mechanism in Adjacent Support Cells. *Cell* 163, 1348–1359. doi: 10.1016/j.cell.2015.10.070

Wang, Y., Sanghvi, M., Gribizis, A., Zhang, Y., Song, L., Morley, B., et al. (2020). Efferent feedback enforces bilateral coupling of spontaneous activity in the developing auditory system. *bioRxiv* [Preprint]. doi: 10.1101/2020.08.12.248799

Wang, H.-Y., Eguchi, K., Yamashita, T., and Takahashi, T. (2020). Frequency-Dependent Block of Excitatory Neurotransmission by Isoflurane via Dual Presynaptic Mechanisms. *J. Neurosci.* 40, 4103–4115. doi: 10.1523/JNEUROSCI.2946-19.2020

Warr, W. B. (1975). Olivocochlear and vestibular efferent neurons of the feline brain stem: their location, morphology and number determined by retrograde axonal transport and acetylcholinesterase histochemistry. *J. Comp. Neurol.* 161, 159–181. doi: 10.1002/cne.901610203

Warr, W. B., and Guinan, J. J. (1979). Efferent innervation of the organ of corti: two separate systems. *Brain Res.* 173, 152–155. doi: 10.1016/0006-8993(79)91104-1

Weisstaub, N., Vetter, D. E., Belén Elgoyhen, A., and Katz, E. (2002). The α9α10 nicotinic acetylcholine receptor is permeable to and is modulated by divalent cations. *Hear. Res.* 167, 122–135. doi: 10.1016/S0378-5955(02)00380-5

Wu, X. S., Sun, J. Y., Evers, A. S., Crowder, M., and Wu, L. G. (2004). Isoflurane inhibits transmitter release and the presynaptic action potential. *Anesthesiology* 100, 663–670. doi: 10.1097/00000542-200403000-00029

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Di Guilmi and Rodríguez-Contreras. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 09 July 2021
doi: 10.3389/fnins.2021.666627

Behavioral Approaches to Study Top-Down Influences on Active Listening

Kameron K. Clayton[1], Meenakshi M. Asokan[1], Yurika Watanabe[1], Kenneth E. Hancock[1,2] and Daniel B. Polley[1,2]*

[1] Eaton-Peabody Laboratories, Massachusetts Eye and Ear, Boston, MA, United States, [2] Department of Otolaryngology – Head and Neck Surgery, Harvard Medical School, Boston, MA, United States

The massive network of descending corticofugal projections has been long-recognized by anatomists, but their functional contributions to sound processing and auditory-guided behaviors remain a mystery. Most efforts to characterize the auditory corticofugal system have been inductive; wherein function is inferred from a few studies employing a wide range of methods to manipulate varying limbs of the descending system in a variety of species and preparations. An alternative approach, which we focus on here, is to first establish auditory-guided behaviors that reflect the contribution of top-down influences on auditory perception. To this end, we postulate that auditory corticofugal systems may contribute to active listening behaviors in which the timing of bottom-up sound cues can be predicted from top-down signals arising from cross-modal cues, temporal integration, or self-initiated movements. Here, we describe a behavioral framework for investigating how auditory perceptual performance is enhanced when subjects can anticipate the timing of upcoming target sounds. Our first paradigm, studied both in human subjects and mice, reports species-specific differences in visually cued expectation of sound onset in a signal-in-noise detection task. A second paradigm performed in mice reveals the benefits of temporal regularity as a perceptual grouping cue when detecting repeating target tones in complex background noise. A final behavioral approach demonstrates significant improvements in frequency discrimination threshold and perceptual sensitivity when auditory targets are presented at a predictable temporal interval following motor self-initiation of the trial. Collectively, these three behavioral approaches identify paradigms to study top-down influences on sound perception that are amenable to head-fixed preparations in genetically tractable animals, where it is possible to monitor and manipulate particular nodes of the descending auditory pathway with unparalleled precision.

Keywords: temporal expectation, top-down, active listening, corticofugal, auditory streaming, descending, efferent, signal detection theory

OPEN ACCESS

Edited by:
Erika Skoe, University of Connecticut, United States

Reviewed by:
Fernando R. Nodal, University of Oxford, United Kingdom
Huizhong Whit Tao, University of Southern California, Los Angeles, United States

****Correspondence:***
Kameron K. Clayton
Kameron_Clayton@meei.harvard.edu

Specialty section:
This article was submitted to Auditory Cognitive Neuroscience, a section of the journal Frontiers in Neuroscience

Received: *10 February 2021*
Accepted: *09 June 2021*
Published: *09 July 2021*

Citation:
Clayton KK, Asokan MM, Watanabe Y, Hancock KE and Polley DB (2021) Behavioral Approaches to Study Top-Down Influences on Active Listening. Front. Neurosci. 15:666627. doi: 10.3389/fnins.2021.666627

INTRODUCTION

During active listening, sound features that are distracting, irrelevant, or totally predictable are often suppressed and do not rise to perceptual awareness (Atiani et al., 2009; Galindo-Leon et al., 2009; O'Connell et al., 2011; Lakatos et al., 2013; Shepard et al., 2016; Sohoglu and Chait, 2016; Southwell et al., 2017). By contrast, anticipated changes in sensory inputs that guide perceptual

decision making are often amplified (Fritz et al., 2003; Gutschalk et al., 2008; Lakatos et al., 2008; Mesgarani and Chang, 2012; Wiegand and Gutschalk, 2012; Atiani et al., 2014; Tsunada et al., 2016; Runyan et al., 2017). For example, a cue that precedes a target stimulus by a fixed temporal interval can provide a salient temporal expectation cue (Nobre and Van Ede, 2018). In active listening paradigms, temporal expectation can increase the probability of sound detection by as much as 40%, a robust effect that has been documented in species ranging from rodents to humans (Wright and Fitzgerald, 2004; Jaramillo and Zador, 2011; Buran et al., 2014; Carcea et al., 2017). The cell types and interconnected circuits that support enhanced processing of anticipated sounds are unknown but is hypothesized to have three essential properties: (i) It would have access to internally generated signals related to timing or preparatory motor actions, (ii) It would connect to lower-level stages of auditory processing to modify the gain and tuning precision of neurons that encode or compute anticipated sound features, (iii) The firing patterns of neurons within this circuit would have to change *before* the onset of an expected sound, to support enhanced processing and perceptual awareness of the subsequent target signal as it arrives.

The massive network of deep layer (L) auditory corticofugal projection neurons fulfill each of the requirements listed above and are therefore a prime candidate for supporting temporally cued active listening. Their apical dendrites reside in superficial cortical layers, where they likely intermingle with inputs from frontal cortex that encode the cue or timing-related inputs initiated by the cue (Xu et al., 2012; Zhang et al., 2014; Takahashi et al., 2016).The axons of deep layer corticofugal neurons innervate subcortical central auditory targets including the medial geniculate body, inferior colliculus, superior olivary complex, and dorsal cochlear nucleus (Diamond et al., 1969; Suga and Ma, 2003; Winer, 2005; Stebbings et al., 2014). As for the final requirement, a recent study from our lab discovered that layer L6 corticothalamic neurons, the largest component of the auditory corticofugal pathway, begin spiking hundreds of milliseconds prior to movements that trigger sounds and rewards (Clayton et al., 2021).

As an example of how corticofugal neurons could mediate perceptual benefits of valid temporal expectations, consider the role of visual cues in the detection of a speech utterance (**Figure 1**). The mouth opens hundreds of milliseconds before speech begins (Chandrasekaran et al., 2009). This visual cue could then be exploited by the auditory system to enhance the processing and intelligibility of a target speech signal amidst a background of competing noise sources (Sumby and Pollack, 1954; Calvert et al., 1997; Grant and Seitz, 2000). In this example and in other types of predictive listening, corticofugal neurons could be the nexus between long-range signals carrying predictive cues and subcortical circuits that process low-level sound features. This hypothesis and others like it could be tested in head-fixed studies of genetically tractable animal models, such as mice, which offer unique advantages over other model systems and freely moving preparations for performing targeted recordings and manipulations of specific types of auditory corticofugal neurons (Bajo et al., 2010; Xiong et al., 2015; Asokan et al., 2017; Guo et al., 2017; Williamson and Polley, 2019; Clayton et al., 2021). However, before launching into the neuroscience studies, a behavioral framework to study temporal expectation in head-fixed mice must first be established.

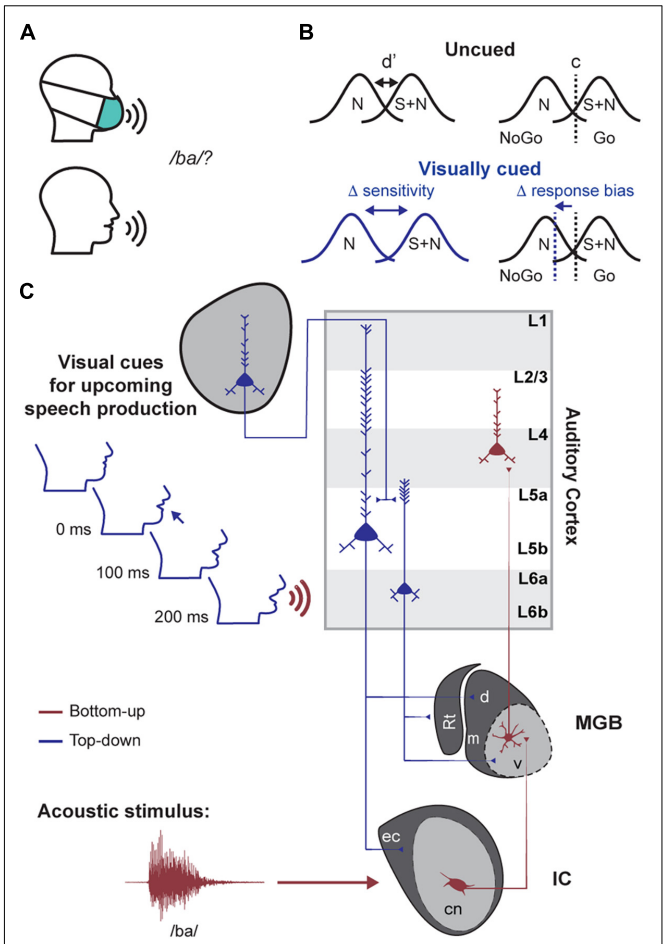

FIGURE 1 | A proposed role for descending projections in predictive listening tasks. **(A)** A basic consonant recognition task is easier in the presence of visual speech cues (e.g., mouth movements). **(B)** Hypothesized changes in signal detection theory metrics between visually cued and uncued conditions. N = noise distribution; S + N = signal and noise response distribution. D-prime (d′) and c refer to the sensitivity and criterion terms, respectively. **(C)** Schematized circuit diagram for one way that cross-modal inputs elicited by facial movement could activate auditory corticofugal neurons for enhanced processing of expected speech cues. Top-down descending auditory corticofugal projections to the inferior colliculus (IC) and medial geniculate body (MGB) are illustrated in blue. Bottom-up corticopetal pathways are illustrated in red. Rt = thalamic reticular nucleus; cn/ec = central nucleus and external cortex of the IC; m/d/v = medial, dorsal and ventral subdivisions of the MGB.

As a first step, this paper details three different operant behavioral approaches to studying predictive listening in head-fixed mice. Across all tasks, we used operant Go/No-Go behaviors that facilitated rapid acquisition of task performance and hundreds of trials per behavioral session, allowing comparison of psychometric functions across conditions where the timing of target stimuli is predictable or not. We interpreted our results through the lens of signal detection theory to tease apart

whether expectation-related changes in perceptual thresholds were mediated by changes in the observer's sensory sensitivity or decision criterion (**Figure 1B**; Green and Swets, 1966; Stanislaw and Todorov, 1999). We provide evidence that sensory and cognitive cues over multiple timescales enhance auditory perception, which provides a behavioral framework for future work that will monitor and manipulate corticofugal neurons during appropriate behaviors to identify their causal involvement in active listening.

MATERIALS AND METHODS

Subjects

All procedures in mice were approved by the Animal Care and Use Committee of the Massachusetts Eye and Ear Infirmary and followed guidelines established by the NIH for the care and use of laboratory animals. A total of 29 mice of both sexes were used. All mice were 6-8 weeks old at the beginning of experiments.

All procedures in humans (N = 9, age range: 20-48, 2 females) were approved by the institutional review board at the Massachusetts Eye and Ear Infirmary. Eligibility of participants was determined by screening for cognitive function (Montreal Cognitive Assessment, MOCA > 25), depression (Beck's depression inventory, BDI < 21 for inclusion), tinnitus (Tinnitus reaction questionnaire, TRQ < 72 for inclusion), use of assistive listening devices (subjects were excluded if they reported use of cochlear implants, hearing aids, bone-anchored hearing aids or FM assistive listening devices) and familiarity with English (subjects were excluded if they did not report at least functional fluency). Normal hearing was confirmed in eligible participants by measuring audiometric thresholds in a double walled acoustic chamber ($\leq$20 dB HL for frequencies up to 8 kHz).

Visually Cued Tone-in-Noise Detection in Human Subjects

Human subject testing occurred in a single-walled walk-in chamber under ambient illumination. Subjects were seated 1 meter in front of a 3 mm red LED and held a response button to indicate detection of the target sound. Sound stimuli were delivered diotically using calibrated headphones (Bose AE2). Digital and analog signals controlling acoustic and visual stimulus presentation were controlled by a PXI system with custom software programmed in LabView (National Instruments). Subjects each completed one 90-min testing session.

In the Go/No-Go tone-in-noise detection task, subjects were required to detect a 12 kHz tone burst (100 ms duration with 5 ms cosine-square onset/offset ramps) of varying intensity in the presence of continuous 50 dB SPL white noise. Subjects indicated detection by button pressing within 2 s following sound onset. Auditory feedback on correct detections was provided with a positive-valence speech token ("Yay!"). No explicit feedback was provided for any other trial outcome (miss, correct reject, or false alarm), but button presses that occurred during catch trials where no tone was presented were followed by a timeout of 5 s. Each trial was followed by an 8-12 s intertrial interval (ITI) drawn from an exponential distribution to maintain a flat hazard rate.

To probe visually cued expectations, we used a 3.5 s visual cue which ramped in brightness in each trial before instantaneously terminating 2 s before tone onset. Performance was compared between interleaved visually cued and uncued trials. To familiarize subjects with the visually cued contingency and determine baseline thresholds, we began by varying target tone level (starting intensity: 40 dB SPL, step size: 1 dB) using a one-up, one-down adaptive procedure with six reversals to determine the subjects' 50% correct thresholds (Levitt, 1971). This adaptive procedure was repeated twice and averaged to determine each subject's detection threshold. In testing blocks (8 blocks of 22 trials each), tone levels for each subject were set relative to the adaptive threshold (−7.5 to 7.5 dB SPL re: threshold in steps of 2.5 dB SPL), with additional catch trials in which no target stimulus was presented. In the test blocks, 33% of trials were visually cued. Tone level and visual cue presentation were randomized across trials.

Preparation for Head-Fixed Mouse Behavior

Prior to behavioral training, mice were implanted with a headplate for subsequent head-fixation during behavioral sessions. Briefly, mice were anesthetized with isoflurane (Piramal) in oxygen (5% induction, 2% maintenance). Lidocaine hydrochloride was administered subcutaneously to numb tissue overlaying the dorsal surface of the skull. The skull was then exposed by retracting the scalp and removing the periosteum. Prior to headplate placement, the skull surface was prepared with etchant (C&B metabond) and 70% ethanol. A custom headplate (iMaterialize, Romero et al., 2020) was then affixed to the skull using dental cement (C&B metabond). Following surgery, Buprenex (0.05 mg/kg) and meloxicam (0.1 mg/kg) were administered and the animal was transferred to a heated recovery chamber.

From 3 to 5 days after the headplate surgery, animals were placed on a water restriction schedule (1 mL/day). Behavioral training began when animals reached a target weight of 80% of their initial body weight. Throughout behavioral training, animals were weighed daily and monitored for signs of dehydration. If mice did not receive 1 mL during a given training session, they were provided with supplemental water. Behavioral sessions took place in dimly lit, single-walled sound-attenuating booths (Acoustic Systems and Med Associates), where mice were placed on an electrically conductive cradle and head-fixed. For tone detection tasks, a single lick spout was positioned 1 cm from the animal's mouth using a 3D micromanipulator (Thorlabs). For the self-initiated frequency recognition task, an apparatus consisting of two lickspouts (4 cm apart) was positioned 1 cm below and 0.5 cm to the right of the animal's snout. Lick spout contact was registered by an electrical circuit which produced a 5 V output signal whenever the tongue closed the circuit between lickspout and the cradle. Lick spout signals were digitized at 400 Hz. Freefield acoustic stimuli were presented through an

inverted dome tweeter positioned 10 cm from the animal's left ear (CUI, CMS0201KLX). A second inverted dome tweeter was placed below the first tweeter at the same distance and azimuthal position for presentation of continuous white noise. Speakers were calibrated before behavioral training using an ultrasonic acoustic sensor (Knowles Acoustics, model SPM0204UD5). For the visually cued tone detection and frequency recognition tasks, a broad spectrum LED (Thorlabs) was placed 20 cm away, in the left visual hemifield. As per human subjects testing, all stimuli, reward delivery, and behavioral contingencies were controlled by a PXI system with custom LabVIEW software.

Behavioral Shaping and Testing in Head-Fixed Mice

Task 1: Light Cued Tone-in-Noise Detection

All mice were habituated to head-fixation for 1-2 sessions before beginning behavioral shaping. Shaping began by conditioning mice to lick for a 70 dB SPL target tone in the presence of 50 dB SPL white noise by presenting a small water reward (4-6 μL) 0.2 s after tone onset. Once licking began to precede reward delivery, animals were moved to an operant version of the task where they were required to lick between 0.2 and 2 s from tone onset to receive reward. Once operant hit rates exceeded 80%, we added additional target intensities to obtain full psychometric functions. Responses during catch trials in which the tone was not presented resulted in a 5 s time out. The visual cue was introduced once detection thresholds reached an asymptote and false alarms rates were consistently below 30% (∼15 sessions into shaping).

Each testing session began by obtaining a 50% detection threshold using a modified 1-up,1-down adaptive procedure (one track with six reversals, 70 dB SPL initial level, 5 dB initial step size, 2.5 dB SPL step size after the first reversal). Catch trials (50% probability) were randomly interleaved in the adaptive track to determine whether psychophysical performance was under stimulus control. Once we had estimated a 50% correct detection threshold, target tone intensities for testing blocks (36 trials, 4-10 blocks per session) spanned −5 dB to +5 dB re: threshold in 2.5 dB steps. In testing blocks, the visual cue was randomly presented in 33% of trials, with an identical waveform and time course as in the human version of the task. Every trial was followed by an 6-10 s ITI drawn from an exponential distribution to maintain a flat hazard rate. Mice performed 200-400 trials per day.

Task 2: Detection of Regular or Jittered Target Streams in a Tone Cloud Background

Mice were maintained on water restriction and adapted to head restraint, as described above. Shaping for the tone-in-cloud detection task was similar to the tone in noise task. In this task, mice were required to detect a repeating 16 kHz target tone (13 individual bursts, each 20 ms in duration, 5 ms cosine-squared ramps) repeated every 480 ms (2.08 Hz) embedded in a continuous tone cloud background. The tone cloud consisted of serially presented tone bursts of varying frequency selected at random (4-48 kHz range of 0.08 octave spacing, 40 dB SPL, 20 ms burst duration, 5 ms onset/offset cosine-squared ramps, 50 Hz repetition rate). A one octave protected bandwidth was included around the target frequency to limit energetic masking (Micheyl et al., 2007; Yin et al., 2007). Each trial began with tone cloud presentation for between 3 and 6 s, randomly drawn from a truncated exponential distribution. If licking occurred in a 2 s window prior to target onset, the countdown to target presentation was extended by another 2 s. Hits were operationally defined as lick spout contact occurring no earlier than 200 ms after the onset of the first target burst and no later than 480 ms after the last burst. Initial conditioning was performed at a signal to noise ratio of 35 dB (75 dB SPL target level). Each trial was followed by an ITI of 4 s. Once animals showed hit rates > 90%, catch trials where no target was presented were introduced. False alarms resulted in a 10 s time out. When false alarm rates fell below 40%, additional target levels were introduced to obtain psychometric functions across a range of signal-to-noise ratios (SNR). Once Go probabilities across catch trials and the full range of SNRs demonstrated that performance was reliably under stimulus control, we introduced a condition where the 13 target tones were presented at a fixed SNR (30 dB) either periodically (at 2.08 Hz) or aperiodically. In the aperiodic condition, the onset timing of tone bursts 2-12 were independently jittered with a time interval selected at random (±20-220 ms). Mice performed 100-200 trials per day.

Task 3: Self-Initiated Frequency Recognition

Mice were maintained on water restriction and adapted to head restraint, as described above. During initial shaping, mice were conditioned to lick the trial initiation spout within 8 s of LED onset to receive a small quantity of water (2 μL). Once mice learned to initiate trials, they were then conditioned to lick the decision spout within 0.2-2 s after the 12 kHz target tone was presented, but not a 6.5 kHz non-target tone (1.5s after initiation, 0.1s tone duration with 5ms raised cosine onset/offset ramps at 70 dB SPL). Lick spout contact during the response window following a non-target (foil) tone resulted in a 4-5 s time out. Contact on the decision spout prior to tone onset ended the trial. Inter-trial intervals were drawn from an exponential distribution (3-10 s). Once mice learned to withhold licking on >80% of foil tones, additional foil frequencies were presented to measure a psychometric function. Once animals displayed <30% false alarm rates for the easiest foil frequency, we introduced blocks of computer-initiated trials where initiation spout licking did not trigger sound. Most mice required 1-2 weeks of behavioral shaping before they could perform the complete frequency recognition task. Daily sessions consisted of 2-5 blocks of self-initiated or computer-initiated trials. Blocks were pseudorandomly interleaved and consisted of 100 trials each. Foil tones were presented in 50% of trials and were randomly selected from five logarithmically spaced frequencies centered on the indecision point (50% false alarm rate) from the previous session. Of 126 behavioral sessions, 10 were excluded either because the mean hit rate was less than 80%, or fewer than 200 total trials were performed.

Once reliable psychometric functions were obtained for self- and computer-initiated trial conditions, we included an addition experiment condition in which the typical 1.5 s foreperiod between self-initiation and target onset was perturbed in a small

fraction (5%) of trials. The particular set of altered foreperiods varied across mice to maximize coverage of a wide-range of delays (0.25 to 1.25 s following self-initiation, in.25 s steps). Violation of the expected foreperiod were always shorter – never longer – than the expected 1.5 s delay.

Data Analysis

Psychometric functions were fit using binary logistic regression. Recognition thresholds (50% Go probability) were determined using the fit psychometric functions. D-prime was calculated as z(hit rate) – z(false alarm rate). The criterion *c* was calculated as – (z(hit rate) + z(false alarm rate))/2 (Stanislaw and Todorov, 1999). For the self-initiated frequency recognition task, d-prime and c values were averaged over all tested frequencies. Across all tasks, reaction time was calculated using the first lick latency on hit trials. Single trial reaction times less than 80 ms were considered artifactual and were not considered for further analysis. For the analysis of perturbations of the self-initiated foreperiod, we z-scored the reaction times from each session with respect to all reaction times for the expected foreperiod. This

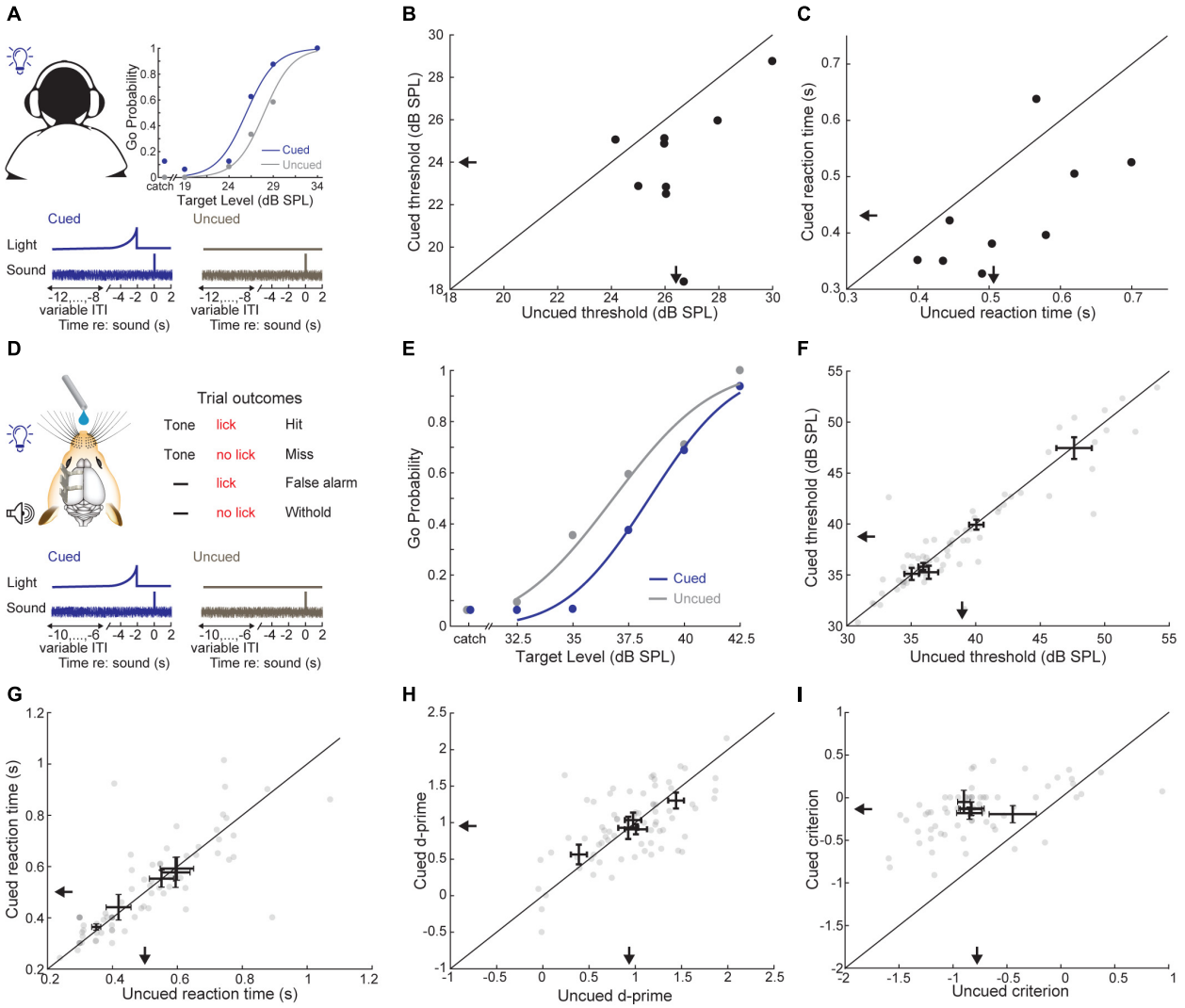

FIGURE 2 | A predictive visual cue decreases tone-in-noise detection thresholds and reaction times in humans, but not mice. **(A)** Schematic of the visually cued tone-in-noise task in human listeners with example psychometric functions for one subject (top) and cue structure (bottom). ITI = inter-trial interval. **(B)** Visually cued and uncued 50% correct thresholds within a behavioral session. Each dot represents an individual subject. Arrows indicate group means. **(C)** Reaction times for visually cued and uncued hit trials. **(D)** Schematic of the visually cued tone-in-noise detection task in water-restricted, head-fixed mice. Trial structure was identical to the human task, except that mice received water rewards on hit trials. **(E)** Probability of a Go response as a function of tone intensity for a single representative behavioral session. Psychometric fits (lines) and actual performance (dots) are shown for visually cued (blue) and uncued trials (gray). In catch trials, the false alarm rate is determined by presenting the visual cue without the auditory target. **(F)** Visually cued and uncued 50% correct thresholds. Each gray dot represents a single behavioral session from one mouse. Bidirectional error bars show mean ± 1 S.E.M. visually cued and uncued thresholds for one mouse. Arrows indicate group means. **(G)** The sensitivity term, d-prime, for visually cued and uncued trials, plotted as in panel **(F)**. **(H)** The criterion term, c, for visually cued and uncued trials, plotted as in panel **(F)**. **(I)** Reaction time (i.e., time to first lick) for visually cued and uncued hit trials, plotted as in panel **(F)**.

approach allowed us to compensate for overall shifts in reaction times across days due to changes in motivation, vigilance, or spout placement.

For all paired difference tests, the mean of each subject's performance across sessions was compared between conditions, as each session from the same mouse could not be considered an independent measurement. Linear regression was used to test foreperiod perturbation effects as each animal was only presented with a pseudorandom subset of delays (mean = 3.25 of 5 possible delays) due to the large number of trials required to obtain psychometric functions for each sparsely presented delay. A linear mixed effects model was used to determine the relationship between frequency recognition thresholds and pre-stimulus licking, while accounting for mouse identity and session numbers as random effects.

RESULTS

Experiment 1: A Predictive Visual Cue Enhances Sound Detection Thresholds in Humans but Not in Mice

To probe the role of predictive sensory cues in auditory perception, we devised a simple cued tone-in-noise detection task where an LED terminated 2 s before tone onset in 33% of trials (**Figure 2A**). First, we attempted to replicate previous studies showing that predictive visual cues enhance auditory detection in humans (Grant and Seitz, 2000) using a Go/No-Go task design that would easily translate to mice. Subjects (N = 9) indicated tone detection by pressing a button and were required to withhold from responding otherwise. We found that tone-in-noise detection thresholds were significantly lower when preceded by the visual cue (cued: 24.03 ± 0.96 dB SPL, uncued: 26.43 ± 0.57 dB SPL, Wilcoxon signed-rank test, $Z = -2.54$, $p = 0.015$; **Figure 2B**). Further, reaction times were shorter for visually cued trials (cued: 0.43 ± 0.03 s, uncued: 0.53 ± 0.03 s, signed-rank test, $Z = -2.31$, $p = 0.02$; **Figure 2C**), consistent with previous work on the role of predictive cueing in sound detection (Greenberg and Larkin, 1968; Wright and Fitzgerald, 2004; Best et al., 2007).

We trained mice in an operant detection task modeled after the conditions used above in human subjects (**Figure 2D**). Across both conditions, mice (N = 5 mice, n = 72 sessions) showed tone detection performance that was qualitatively similar to human observers, with low false alarm rates on catch trials and steeply sloping psychometric functions across a range of target tone intensities (**Figure 2E**). However, we did not find any consistent benefit for the visual cue on detection thresholds (cued: 38.71 ± 2.37 dB SPL, uncued: 39.00 ± 2.32 dB SPL, signed-rank test, $Z = -1.75$, $p = 0.08$; **Figure 2F**) or reaction times (cued: 0.50 ± 0.04 s, uncued: 0.50 ± 0.05 s, signed-rank test, $Z = 0.40$, $p = 0.69$; **Figure 2G**). To better understand if the weak effects of the visual cue on threshold and reaction time belied underlying changes in sensitivity or response bias, we turned to signal detection theory, which provides a means for formal assessment of both measures (Green and Swets, 1966; Stanislaw and Todorov, 1999). Visual cueing had little effect on the separability of signal and noise distributions, as measured by d-prime (cued: 0.95 ± 0.12 stds, uncued: 0.95 ± 0.17 stds, signed-rank test, $Z = 0.13$, $p = 0.89$; **Figure 2H**). By contrast, the response bias was significantly reduced (i.e., biased toward NoGo rather than Go responses) when target tones were preceded by a visual cue (cued: −0.14 ± 0.03 stds, uncued: −0.77 ± 0.08 stds, signed-rank test, $Z = 2.02$, $p = 0.04$; **Figure 2I**). These findings suggest that the visual cue was perceived by both species. Human subjects were able to exploit the visual predictive cue to more reliably perceive liminal tones in noise, whereas in mice, the visual cue altered their overall behavioral response bias without affecting their perceptual sensitivity to the target stimulus.

Experiment 2: Mice Exploit Temporal Regularities to Perform an Auditory Streaming Task

Interactions between visual predictive cues and auditory targets could be constrained by multi-sensory temporal binding windows derived from natural scene statistics or neural circuits subserving multisensory integration (van Wassenhove et al., 2007; Stevenson et al., 2011). In mice, the 2 s delay between visual cue offset and sound onset may have proved too long a gap for the visual stimulus offset to facilitate auditory detection (Siemann et al., 2015). To study

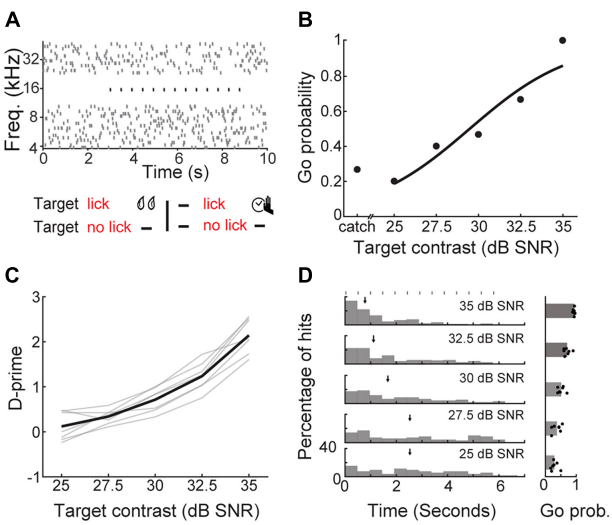

FIGURE 3 | Detection of a periodic target in background noise. **(A)** Schematic of a Go/No-Go detection task that required mice to report the perception of a repeating target tone in the presence of a tone cloud background. Top; the stimulus spectrogram for a trial in which the repeated target tone (black dots) begins 3 s after the onset of the random, serially presented masking tones (gray dots). Bottom; task contingencies. **(B)** Go probability as a function of target contrast for a single representative session. **(C)** D-prime as a function of target contrast, showing the mean (black line) and individual mice (gray lines). **(D)** Left, reaction time distributions as a function of target contrast. Median reaction times are indicated with arrows. Target burst timing is indicated by gray rectangles. Right, Go probability as a function of target contrast, showing the mean and individual values (dots).

the perceptual benefits of auditory temporal expectation in mice without relying on cross-modal integration, we turned to a within-modality cue. In auditory scene analysis, repetition provides a salient grouping cue for separating an auditory stream or object from background stimuli (Kidd et al., 1994; Gutschalk et al., 2008; Agus et al., 2010; Andreou et al., 2011; McDermott et al., 2011). In human listeners, presenting a repeated signal in the midst of an ongoing mixture produces a stream segregation phenomenon where the repeated signal pops out from the mixture after several repetitions (Micheyl et al., 2007; Gutschalk et al., 2008; McDermott et al., 2011). We tested whether a similar phenomenon existed in mice by tasking them with operantly reporting the presence of a regularly repeated target tone embedded within a tone cloud, with similar outcome contingencies as in the visual-cued tone detection task (**Figure 3A**).

To determine if mice could perform stream segregation using embedded repetition, we varied the sound level of the target tone relative to the tone cloud to obtain psychometric functions (N = 7 mice, n = 18 sessions; **Figure 3B**). As the signal to noise ratio (SNR) improved, d-prime increased (one-way repeated measures ANOVA, $F(6, 24) = 152.17$, $p < 1 \times 10^{-15}$; **Figure 3C**). At high SNRs, mice typically responded after the first target tone burst, suggesting that the intensity contrast between the target and the background was sufficient to support detection (**Figure 3D**). At less favorable SNRs, median reaction times occurred after two to three target tone bursts, consistent with the time course of build-up for repeating targets in human listeners. However, the longer reaction times we observed with decreasing SNR could also be attributed to the reduced stimulus intensity and not the regular repetition of the target, *per se* (Piéron, 1913; Viemeister and Wakefield, 1991).

To directly test whether long reaction times were a result of stream segregation based on rhythmic repetition, we jittered the repetition rate of the target tone at an intermediate SNR (30 dB), where reaction times suggested that target perception might benefit from temporal integration. We reasoned that if detection reflected a purely probabilistic process based on the stimulus contrast for each individual tone, presenting the targets aperiodically would not impact detection probability or reaction time distributions (**Figure 4A**). However, if target recognition benefited from the regular periodicity of the target, the likelihood of target detection would increase with repetitions of the target, and reaction times for regular targets would be skewed toward later repetitions.

Consistent with the latter hypothesis for temporal integration, mice were more likely to detect the target stream when it was regularly repeated (N = 6 mice, n = 19 sessions; periodic: 56.14 ± 2.10%, aperiodic: 43.34 ± 3.52%, signed-rank test, $Z = 2.20$ $p = 0.03$; **Figure 4B**). More to the point, median reaction times were significantly longer on hit trials in the periodic condition compared to the aperiodic condition (periodic: 4.51 bursts or 2.16 s, aperiodic: 2.37 bursts or 1.14 s, signed-rank test, Z = 2.21, p = 0.03; **Figure 4C**), suggesting that the improved overall detection probability with periodic target rates could be attributed to an increased probability of target detection later in the stream, which did not occur in the aperiodic condition. To determine if the distribution of reaction times between regular and jittered trials was different, we pooled data across all subjects and sessions. While the modal reaction time in hit trials corresponded to detection of the first tone in the target, over 34.67% of regular hits occurred after 6 bursts, compared to only 17.42% of the jittered hits (Kolmogorov-Smirnov two-tailed test, D = 0.23, p = 0.00004; **Figure 4D**).

Experiment 3: Self-Initiation Enhances Frequency Discrimination

While each of the paradigms described above demonstrate the use of top-down cues in listening tasks, neither is optimized for future studies that combine behavioral assessments in head-fixed mice with neurophysiological methods to selectively monitor and manipulate auditory corticofugal neurons. For studies that will isolate the causal involvement of corticofugal neurons in temporal expectation, a behaviorally quiescent period between the cue and target sound, during which no explicit input is provided could prove useful for homing in on neural preparatory activity (Buran et al., 2014; Carcea et al., 2017). Further, if increased neural activity was observed prior to the onset of the target sound in either of the

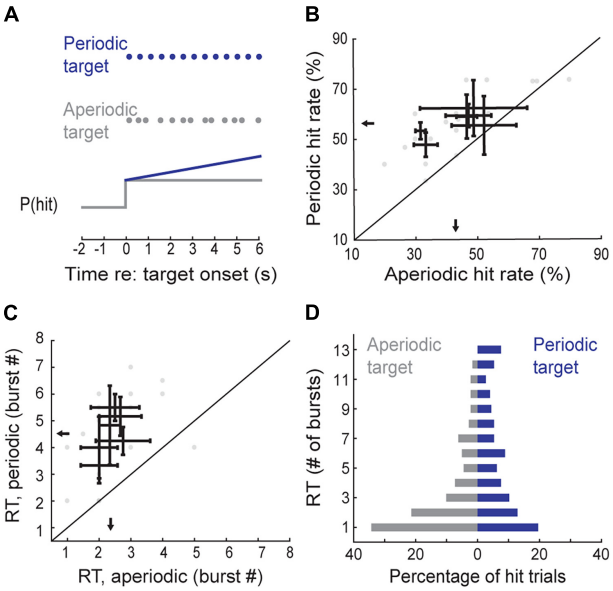

FIGURE 4 | Temporal regularity enhances the perceptual salience of a repeated target. **(A)** *Top*, schematic of periodic and aperiodic target presentation. *Bottom*, hypothesized "build-up" of detection probability based on temporal regularity. **(B)** Hit rates for interleaved periodic and aperiodic targets presented at a fixed contrast (30 dB SNR). Data from individual sessions are shown as gray circles. Black lines represent the mean ± SEM hit rates for each mouse. Arrows indicate group means. **(C)** Reaction times, in number of bursts, plotted as in panel **(B)**. **(D)** Histograms of reaction times, pooled across all session and subjects, for periodic and aperiodic trains of target bursts.

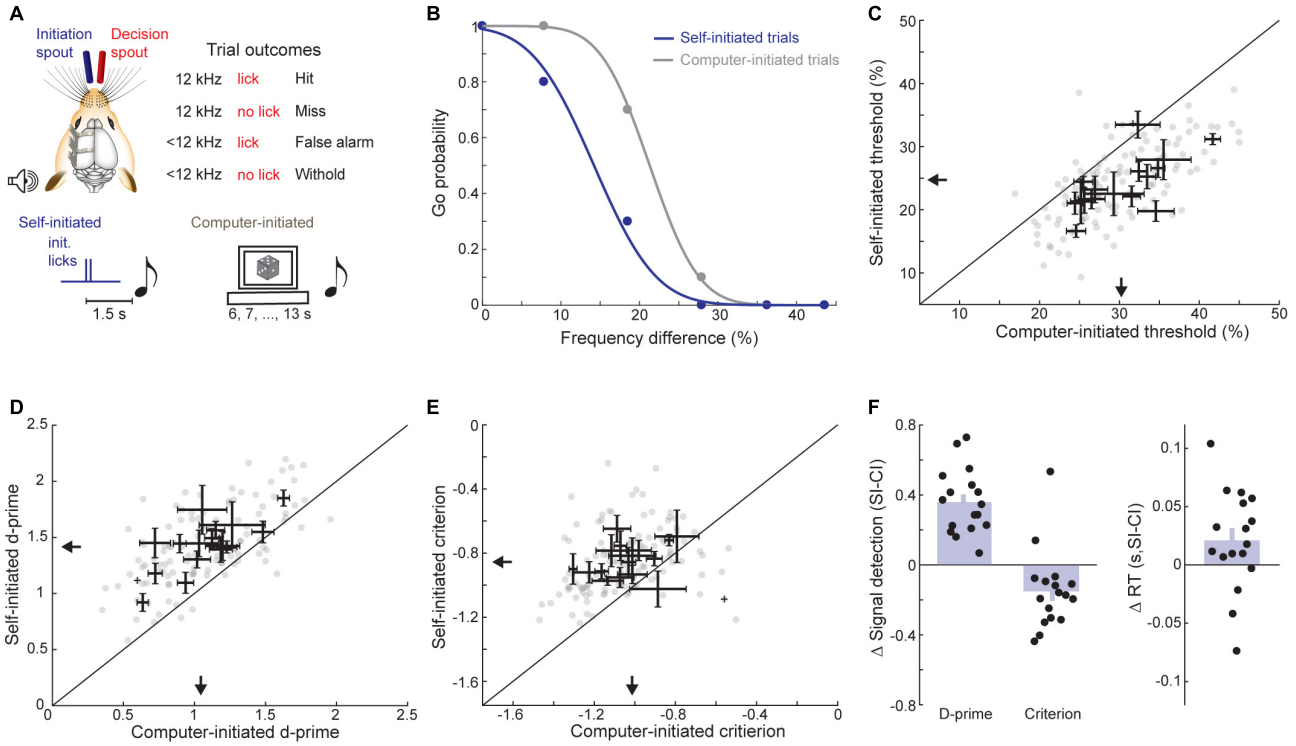

FIGURE 5 | Trial self-initiation enhances frequency recognition thresholds. **(A)** Task design. *Top*, schematic of task set-up and contingencies for decision spout licking. *Bottom*, Target (12 kHz) and foil tones (<12 kHz) are presented with predictable timing following contact with the initiation spout (self-initiated, left) or with unpredictable timing determined by the computer (computer-initiated, right). **(B)** Representative psychometric functions from a single session with interleaved self- and computer-initiated trial blocks. **(C)** Self-initiated and computer-initiated 50% correct thresholds for frequency recognition. Each gray dot represents one behavioral session. Bidirectional error bars show mean ± 1 S.E.M. self-initiated and computer-initiated thresholds for each mouse. Arrows indicate group medians. **(D)** D-prime value, plotted as in panel **(C)**. **(E)** Criterion values, plotted as in panel **(C)**. **(F)** Quantification of mean changes in sensitivity, criterion, and reaction time between self-initiated and computer-initiated trials. Each black dot represents one subject and bars reflect the mean ± 1 SEM.

temporal expectation behaviors presented above, it would be difficult to disambiguate whether this activity reflected a neural representation of expectation or a motor preparatory signal related to the impending Go response (Clayton et al., 2021).This ambiguity could be resolved by task designs that either require subjects to make an overt behavioral report in all trials (e.g., alternative forced choice) or task designs where animals can make a correct decision by withholding a behavioral report.

To address these limitations, we turned to a frequency recognition task where mice were trained to lick a decision spout following a 12 kHz tone, but withhold licking for tones of other frequencies (**Figure 5A**, top). To manipulate the availability of top-down information related to stimulus timing, target stimuli were occasionally presented at a fixed interval following self-initiated movement, rather than a cross-modal sensory cue, as it rules out the possibility that the preparatory cue was not detected (Reznik et al., 2014; Morillon et al., 2015). Trial self-initiation is a routine component of non-human primate and freely moving rodent behavioral tasks (Ghose and Maunsell, 2002; Nakajima et al., 2019). In freely moving rats and gerbils, sound detection and discrimination thresholds are better when animals self-initiate trials compared to conditions where sound presentation is unpredictable (Jaramillo and Zador, 2011; Buran et al., 2014; Carcea et al., 2017).

In our frequency recognition task, mice (N = 17 mice, n = 115 sessions) triggered sound 1.5 s later by licking a separate initiation spout to receive a small reward (**Figure 5A**, bottom). We contrasted blocks of self-initiated trials with blocks of computer-initiated trials where sound presentation timing was unrelated to contacts on the self-initiation spout. Psychophysical performance was under stimulus control in both conditions, as evidenced by high hit rates to the 12 kHz target and declining false positive rates at foil frequencies further away from the target (**Figure 5B**). Importantly, a clear top-down influence was observed in this behavior, as a clear reduction in target recognition threshold was observed across all mice and virtually all sessions in self-initiated trials (self-initiated: 24.61 ± 1.16%, computer-initiated: 30.35 ± 1.19%, sign-rank test, Z = −3.38, p = 0.0007, **Figure 5C**). Improved thresholds for self-initiated trials were mediated by an increased d-prime (self-initiated: 1.41 ± 0.06 stds, computer-initiated: 1.05 ± 0.07 stds, signed-rank test, Z = 3.62, p = 0.0003; **Figures 5D,F**) and a decreased

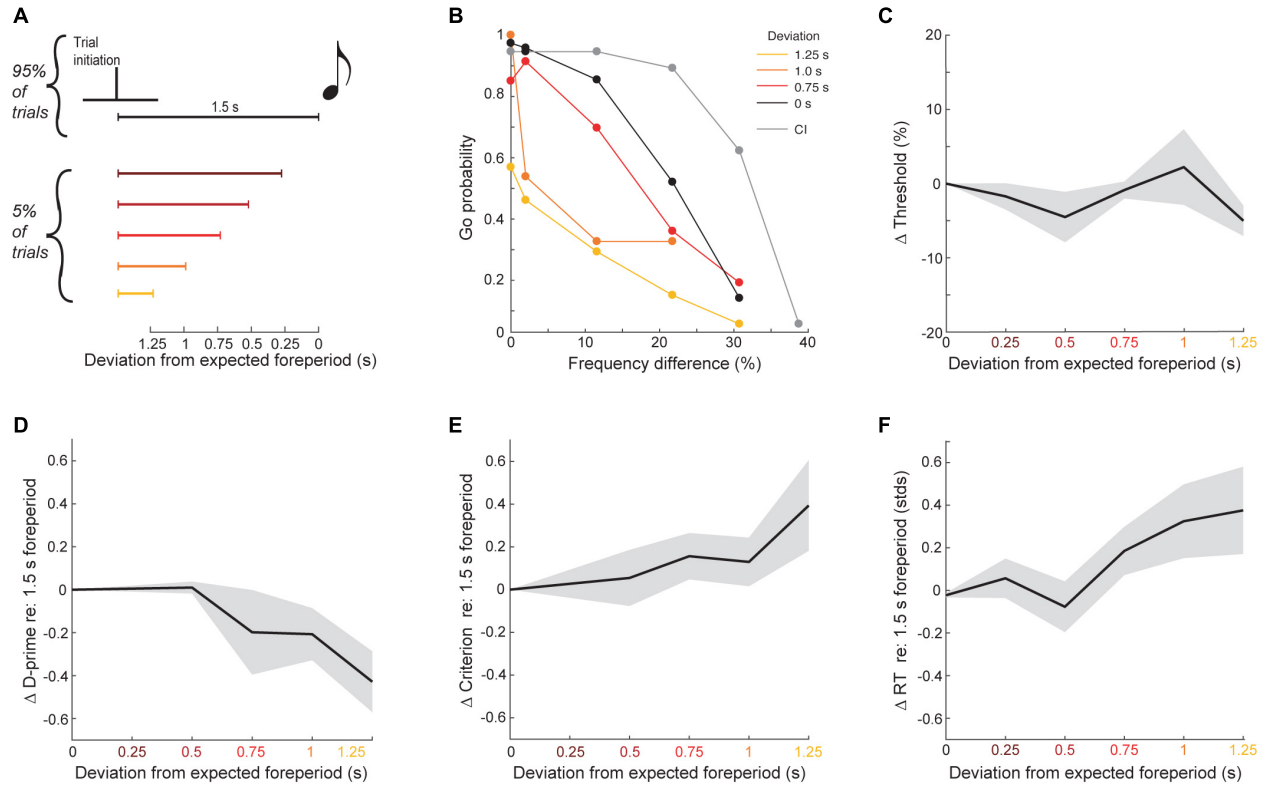

FIGURE 6 | Frequency recognition is impaired when tones occur at unexpected times. **(A)** Schematic of foreperiod violations. **(B)** Go probability as a function of frequency difference for each tested delay in one example mouse. **(C)** Mean ± SEM change in thresholds as a function of deviation from the expected foreperiod across all mice. **(D)** Change in d-prime, as in panel **(C)**. **(E)** Change in criterion, as in panel **(C)**. **(F)** Change in z-scored reaction times, as in panel **(C)**.

bias to respond (self-initiated criterion: −0.86 ± 0.02 stds, computer-initiated: −1.01 ± 0.18 stds, signed-rank test, $Z = 2.53$, $p = 0.01$, **Figures 5E,F**). Contrary to previous studies in rats (Jaramillo and Zador, 2011; Carcea et al., 2017), we observed a tendency for slower reaction times on self-initiated trials (self-initiated: 0.24 ± 0.01 s, computer initiated: 0.22 ± 0.01 s; signed-rank test, $Z = 1.96$, $p = 0.049$; **Figure 5F**). Differences between this result and previous studies could be due to species differences or, more likely, to differences in the operant behavior (i.e., between freely moving nose-poke versus head-fixed licking).

Sparse Violations of Temporal Expectations Reveal a Temporal Filter for Optimal Performance

The perceptual benefit of self-initiation could reflect the specific benefit of temporal expectations or simply a non-specific increase in arousal during blocks of self-initiated trials. To test whether the top-down benefit on self-initiated trials were consistent with a narrow window of increased temporal expectation or a more global elevation of arousal, we changed the foreperiod separating contact on the initial spout and sound onset from the typical 1.5 s duration to a shorter interval on a small fraction of trial (5%, $N = 12$ mice, $n = 37{,}578$ total trials over 200 sessions; **Figure 6A**). Violating the typical 1.5 s foreperiod by hundreds of milliseconds had striking effects on task performance, suggesting that self-initiation benefits reflected a specific time window of expected stimulus arrival. As shown in an example mouse, the typical self-initiation benefit seen on the majority of trials (black vs. gray lines in **Figure 6B**), was progressively changed when target and foil sounds were presented earlier than the expected timing. Importantly, the effect of foreperiod perturbations could not be seen in simple reports of frequency recognition threshold, on account of changes in both the hit and false positive rates (linear regression, $F(1,48) = 1.29$, $p = 0.26$, $R^2 = 0.03$; **Figure 6C**). A signal detection theory analysis found that d-prime decreased as the sound onset deviated more from the expected timing, indicating that mice were less able to discriminate between target and foil at unexpected intervals (linear regression, $F(1,48) = 6.74$, $p = 0.02$, $R^2 = 0.21$; **Figure 6D**). The criterion also became more positive, indicating that mice were less likely to respond at greater foreperiod violations and – based on the increased probability of misses to the 12 kHz target stimulus – may not have been perceptually aware that the stimulus was presented at all (linear regression, $F(1,48) = 5.09$, $p = 0.03$, $R^2 = 0.16$; **Figure 6E**). Consistent with a specific temporal filter for listening, violations of the expected foreperiod duration were also associated with increased reaction times on Go trials (linear regression, $F(1,48) = 7.29$, $p = 0.009$, $R^2 = 0.13$; **Figure 6F**).

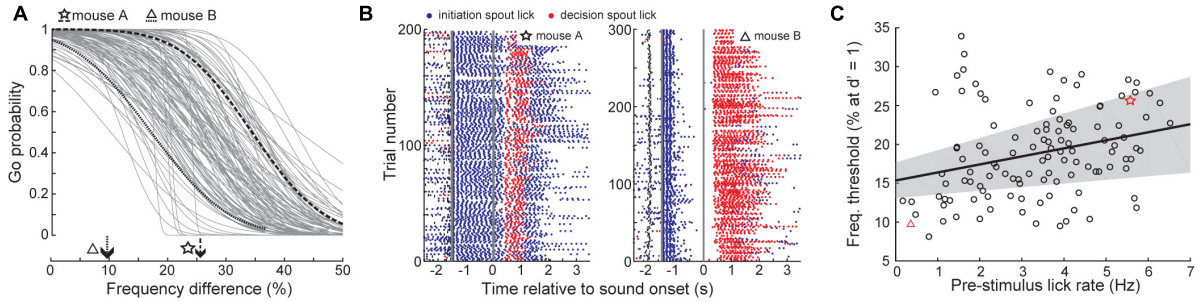

FIGURE 7 | Vigorous licking during the foreperiod is associated with impaired frequency discrimination. **(A)** Psychometric functions for self-initiated frequency recognition trials across all behavioral sessions. **(B)** Lick rasters for the two example sessions shown in panel **(A)**. Blue and red dots denote lick timing on the initiation and decision spouts, respectively. Thick gray vertical lines indicate the timing of trial initiation and sound onset. Small vertical black lines indicate the timing of visual ready cue. **(C)** Frequency thresholds are improved in sessions with minimal licking prior to tone onset. Thick black line and gray shading represent the linear regression line and the bootstrapped 95% confidence interval of the linear regression, respectively. Each individual test session is represented by a dot, red symbols indicate example behavior sessions shown in panels **(A,B)**.

Frequency Discrimination Thresholds Are Influenced by How Animals Initiate Trials

While reaction times were slower at unexpected intervals after self-initiation, it was surprising that self-initiated reaction times were 20 ms longer than for computer-initiated trials, as valid temporal expectations often speed decisions (Jaramillo and Zador, 2011). Further, we observed substantial heterogeneity in self-initiated thresholds across sessions (**Figure 7A**). To explain these outstanding sources of variability, we examined how mice performed the task at a more granular level by quantifying licking activity on the trial initiation spout. We noted that behavioral sessions with poor discrimination thresholds were associated with persistent licking of the initiation spout throughout the foreperiod (**Figure 7B**, left). In sessions with better discrimination thresholds, mice briefly licked the initiation spout and then paused before the target or non-target tone was presented (**Figure 7B**, right). Across all behavioral sessions, initiation spout lick rate during the foreperiod was negatively correlated with the discrimination threshold, even after controlling for other mitigating variables including the individual mouse and the number of prior testing sessions (linear mixed effects model, slope = −1.11% freq. diff/lick, $t(117) = 2.69$, $p = 0.008$; **Figure 7C**).

DISCUSSION

Knowing when to listen can enhance the detection of faint sounds or the discrimination of target sounds from distractors (Egan et al., 1961; Greenberg and Larkin, 1968; Wright and Fitzgerald, 2004; Best et al., 2007; Nobre and Van Ede, 2018). Here, we reported three inter-related behavioral experiments that explored how predictive sensory and motor cues enhance the perceptual detection and discrimination of upcoming sounds in humans and mice. Depending on the modality and timing of the predictive cue, valid expectations altered the observer's perceptual sensitivity, response bias, or both. We found that a visual cue provided significant perceptual benefit to human listeners in a tone-in-noise detection task, but only a response criterion change in mice (**Figure 2**). A second experiment found periodicity aided the detectability of a liminal repeating target in a complex background noise by increasing the probability of late detection events, after the regularity of the target sound had been established (**Figures 3**, **4**). In a third paradigm, we trained mice to expect a tone stimulus at a fixed interval after motor self-initiation, which results in significantly lower frequency recognition thresholds and improved sensitivity relative to unpredictable computer-initiated trials (**Figure 5**). The improved perceptual sensitivity for self-initiated sound was abolished when sound arrived approximately 1 s earlier than expected, suggesting a specific temporal window for enhanced auditory perception (**Figure 6**). A closer inspection of self-initiated frequency recognition trials revealed that exuberant motor activity during the foreperiod also interfered with frequency recognition accuracy, suggesting different types of internally generated signals that enhance or degrade perceptual performance (**Figure 7**).

Our approach to studying perceptual expectations in head-fixed mice was inspired by previous human psychophysical studies (Wright and Fitzgerald, 2004; Best et al., 2007). While head-fixed mice can perform two-alternative forced choice tasks (Sanders and Kepecs, 2012; Burgess et al., 2017; Vincis et al., 2020), we used Go/No-Go task designs which could be learned over the course of one to two weeks, allowing us to layer manipulations of perceptual expectations on top of basic detection or discrimination tasks (Guo et al., 2014; McGinley et al., 2015; Kuchibhotla et al., 2017). As appetitive Go/No-Go behaviors typically result in a bias toward responding (Gomez et al., 2007), the role of baseline response bias and motivational structure introduced by our task designs likely played a role in the expectation-related effects we observed. For instance, despite human listeners showing lower thresholds and faster responses in a visually cued tone-in-noise detection task, there was no effect of visual cueing on mouse detection thresholds or reaction times. Instead, we observed a shift in criterion in mice that led them

to respond less frequently during visually cued trials across all sound levels. Similarly, in the self-initiated frequency recognition task, we observed decreased bias to respond during self-initiated trials, though the change in bias reported in this task were substantially smaller than the change in d-prime. The Go/No-Go tasks presented here have asymmetrical response requirements: while Go responses require just a single lick, No-Go responses require mice to withhold from licking throughout the entire response period. In this response structure, self-initiation or visual cueing could provide a warning signal to temporarily withhold from licking unless there is strong evidence that the target sound is present. Though the generality of our findings await testing in other operant behavioral task designs, our results demonstrate that predictive cues have varied effects on response bias and sensitivity that should be disambiguated by using signal detection theory or similar techniques.

Beyond the influence of task structure and reward contingencies, our results suggest that the effects of temporal expectations in head-fixed mice vary according to the modality and timing of predictive cues and the sounds they predict. In the visually cued task, the lack of changes in threshold or sensitivity we found might have been attributable to the relatively long delay between the visual cue and auditory target onset, as suggested from a recent report demonstrating improved tone in noise discrimination using a more proximal visual cue (400 ms foreperiod, Nakajima et al., 2019). Further, the spatial position of the visual cue could have contributed to our results in mice, a more ethologically relevant spatial position may have increased its behavioral salience (e.g., Yilmaz and Meister, 2013). Our second paradigm focused on intra-modal cues that can aid target perception on fast timescales. We found that SNR and temporal regularity provided independent bottom-up and top-down cues, respectively, to identify a target in a mixture. Our results in mice are consistent with previous studies in human listeners which collectively demonstrated that repetition is a salient grouping cue for auditory scene analysis (Kidd et al., 1994; Micheyl et al., 2007; Agus et al., 2010; Andreou et al., 2011; McDermott et al., 2011). The prolonged time course of repetition-based stream segregation suggests a mechanism by which repetitive inputs are integrated over time and used to predict the incoming acoustic signal. Our work provides behavioral proof-of-principle for future studies to uncover the neural substrates of this prolonged temporal integration process in a genetically tractable model organism.

Our third paradigm confirmed prior reports of improved thresholds and perceptual sensitivity when target sounds are presented at fixed time intervals following movement-based trial self-initiation (Jaramillo and Zador, 2011; Buran et al., 2014; Carcea et al., 2017). Improved discrimination in self-initiated trials could reflect differences in arousal or other global internal state differences (Rodgers and DeWeese, 2014; McGinley et al., 2015). By perturbing trial timing in a small fraction of trials, we showed that perceptual sensitivity decreased when sounds were presented at unexpected moments, arguing against purely arousal-mediated changes between self-initiated and computer-initiated trials. The pathways and processes underlying perceptual changes in self-initiated trials are less clear. Self-initiated movements can directly modulate central auditory processing, either through internal motor-corollary inputs or reafferent sensory inputs (Reznik et al., 2018; Schneider et al., 2018). Our findings in the self-initiated task suggest the combination of two opposing top-down influences on sound perception: on the one-hand temporal expectation clearly improved frequency recognition thresholds and perceptual sensitivity (**Figure 5**). At the same time, when movement-based contact on the trial initiation spout impinged too closely upon the sound delivery period, it adversely affected frequency recognition thresholds.

The generation of temporal expectations likely involves distributed circuits across frontal and parietal cortical areas, basal ganglia, hippocampus, and cerebellum (Janssen and Shadlen, 2005; MacDonald et al., 2013; Narain et al., 2018; Zhou et al., 2020). Many of the same regions implicated in temporal expectations are sensory responsive, but sound representations in these areas are context-dependent and evolve at slower timescales compared to representations within the early central auditory pathway (Rummell et al., 2016; Runyan et al., 2017; Elgueda et al., 2019). Our behavioral data show that top-down influences related to temporal expectations enhance even basic perceptual abilities like tone detection or recognition. Changes in basic auditory processing driven by top-down influences suggest a scheme where long-range inputs from brain areas involved in generating temporal expectations would modulate the fast-timescale, fine-grained encoding of acoustic features, which is generally restricted to the beginning of the central auditory pathway (Joris et al., 2004; Asokan et al., 2021).

While previous studies have reported changes in pre-stimulus neural activity during self-initiated auditory perceptual tasks, the precise neural circuits responsible for transforming long-range predictive inputs into changes in local network excitability based on temporal expectations and motor-corollary inputs remain relatively unexplored. Multiple lines of evidence from our lab suggest that a particular subclass of auditory corticofugal projection neuron, the layer 6 corticothalamic neuron (L6 CT), may play a central role in this process. First, artificial activation of auditory L6 CTs can enhance or impair sound discrimination, depending on the temporal delay between L6 CT spiking and sound presentation (Guo et al., 2017). At short delays between optogenetically induced L6 CT spiking and sound presentation (i.e., 100 ms), behavioral sound discrimination is enhanced, while at long delays (i.e., 200 ms), sound discrimination is impaired. Second, L6 CTs receive direct long-range inputs from motor-related areas such as the globus pallidus and increase their spiking hundreds of milliseconds before movements which predict sound and reward (Clayton et al., 2021). During trial self-initiation, L6 CTs would presumably be activated prior to contact with the initiation spout, shifting auditory cortex network excitability into an optimal state for discrimination, consistent with the improved behavioral discrimination we observed. However, any benefit derived from L6 CT activation would also depend on the precise alignment between L6 CT spiking and sound presentation, which may account for our observation that initiation spout licking which impinged on the sound delivery period impaired frequency recognition thresholds. While self-initiation is a strong

predictive cue, future work which decouples movement-related activity in L6 CTs from cues that predict sound presentation timing could better elucidate how long-range predictive inputs sculpt L6 CT spiking to shift the auditory cortical network into an optimal state to process sounds at expected moments according to behavioral goals.

A role of auditory corticofugal cell-types in auditory processing and perception will likely require careful analysis of targeted corticofugal cell types in behaving animals. Although few studies have specifically investigated corticofugal contributions to sound-guided behaviors, their findings highlight a critical role of descending projections in contextual processing and experience-dependent plasticity (Bajo et al., 2010; Xiong et al., 2015; Guo et al., 2017; Homma et al., 2017). Seminal cell-type specific ablation studies found that auditory cortex neurons which project to the inferior colliculus play a key role in sensorimotor remapping after monaural deprivation (Bajo et al., 2010). Other work has shown that corticollicular neurons also control sound-driven innate defensive behaviors such as escape (Xiong et al., 2015). However, despite the potential role of descending projections in real-time subcortical gain control according to behavioral goals or attention, the necessary involvement of corticofugal neurons in these behaviors have yet to be demonstrated. With the development of predictive listening paradigms in head-fixed mice described here, a more complete understanding of how descending control in the auditory pathway guides adaptive behavior and active listening is within reach.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by The Institutional Review Board at the Massachusetts Eye and Ear Infirmary. The patients/participants provided their written informed consent to participate in this study. The animal study was reviewed and approved by Animal Care and Use Committee of the Massachusetts Eye and Ear Infirmary.

AUTHOR CONTRIBUTIONS

KC, MA, and DP designed the study. KC, MA, and YW performed all the experiments and analyzed the data with supervisory input from DP. KH programmed the behavioral data acquisition software. KC and DP wrote the manuscript with feedback from all authors. All authors contributed to the article and approved the submitted version.

FUNDING

Experiments using mice as experimental subjects was supported by the NIH grant DC017178 (DP) and NSF fellowship DGE1745303 (KC). Experiments using human subjects was supported by the NIH grant DC015857 (DP).

ACKNOWLEDGMENTS

We thank C. Liu for assistance with behavioral training. We also thank E. Foss for supporting the design and fabrication of electronic components used in mouse behavioral tasks.

REFERENCES

Agus, T. R., Thorpe, S. J., and Pressnitzer, D. (2010). Rapid formation of robust auditory memories: insights from noise. *Neuron* 66, 610–618. doi: 10.1016/j.neuron.2010.04.014

Andreou, L. V., Kashino, M., and Chait, M. (2011). The role of temporal regularity in auditory segregation. *Hear. Res.* 280, 228–235. doi: 10.1016/j.heares.2011.06.001

Asokan, M. M., Williamson, R. S., Hancock, K. E., and Polley, D. B. (2017). Homeostatic normalization of sensory gain in auditory corticofugal feedback neurons. *bioRxiv* [Preprint]. doi: 10.1101/162909

Asokan, M. M., Williamson, R. S., Hancock, K. E., and Polley, D. B. (2021). Inverted central auditory hierarchies for encoding local intervals and global temporal patterns. *Curr. Biol.* 31, 1762.e4–1770.e4. doi: 10.1016/j.cub.2021.01.076

Atiani, S., David, S. V., Elgueda, D., Locastro, M., Radtke-Schuller, S., Shamma, S. A., et al. (2014). Emergent selectivity for task-relevant stimuli in higher-order auditory cortex. *Neuron* 82, 486–499. doi: 10.1016/j.neuron.2014.02.029

Atiani, S., Elhilali, M., David, S. V., Fritz, J. B., and Shamma, S. A. (2009). Task difficulty and performance induce diverse adaptive patterns in gain and shape of primary auditory cortical receptive fields. *Neuron* 61, 467–480. doi: 10.1016/j.neuron.2008.12.027

Bajo, V. M., Nodal, F. R., Moore, D. R., and King, A. J. (2010). The descending corticocollicular pathway mediates learning-induced auditory plasticity. *Nat. Neurosci.* 13, 253–260. doi: 10.1038/nn.2466

Best, V., Ozmeral, E. J., and Shinn-Cunningham, B. G. (2007). Visually-guided attention enhances target identification in a complex auditory scene. *J. Assoc. Res. Otolaryngol.* 8, 294–304. doi: 10.1007/s10162-007-0073-z

Buran, B. N., von Trapp, G., and Sanes, D. H. (2014). Behaviorally gated reduction of spontaneous discharge can improve detection thresholds in auditory cortex. *J. Neurosci.* 34, 4076–4081. doi: 10.1523/jneurosci.4825-13.2014

Burgess, C. P., Lak, A., Steinmetz, N. A., Zatka-Haas, P., Bai Reddy, C., Jacobs, E. A. K., et al. (2017). High-yield methods for accurate two-alternative visual psychophysics in head-fixed mice. *Cell Rep.* 20, 2513–2524. doi: 10.1016/j.celrep.2017.08.047

Calvert, G. A., Bullmore, E. T., Brammer, M. J., Campbell, R., Williams, S. C. R., McGuire, P. K., et al. (1997). Activation of auditory cortex during silent lipreading. *Science* 276, 593–596. doi: 10.1126/science.276.5312.593

Carcea, A. I., Insanally, M. N., and Froemke, R. C. (2017). Dynamics of auditory cortical activity during behavioral engagement and auditory perception. *Nat. Commun.* 8, 1–12.

Chandrasekaran, C., Trubanova, A., Stillittano, S., Caplier, A., and Ghazanfar, A. A. (2009).). The natural statistics of audiovisual speech. *PLoS Comput. Biol.* 5:e1000436. doi: 10.1371/journal.pcbi.1000436

Clayton, K. K., Williamson, R. S., Hancock, K. E., Tasaka, G., Mizrahi, A., Hackett, T. A., et al. (2021). Auditory corticothalamic neurons are recruited by motor preparatory inputs. *Curr. Biol.* 31, 310.e5–321.e5.

Diamond, I. T., Jones, E. G., and Powell, T. P. S. (1969). The projection of the auditory cortex upon the diencephalon and brain stem in the cat. *Brain Res.* 15, 305–340. doi: 10.1016/0006-8993(69)90160-7

Egan, J. P., Greenberg, G. Z., and Schulman, A. I. (1961). Interval of time uncertainty in auditory detection. *J. Acoust. Soc. Am.* 33, 771–778.

Elgueda, D., Duque, D., Radtke-Schuller, S., Yin, P., David, S. V., Shamma, S. A., et al. (2019). State-dependent encoding of sound and behavioral meaning in a tertiary region of the ferret auditory cortex. *Nat. Neurosci.* 22, 447–459. doi: 10.1038/s41593-018-0317-8

Fritz, J., Shamma, S., Elhilali, M., and Klein, D. (2003). Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex. *Nat. Neurosci.* 6, 1216–1223. doi: 10.1038/nn1141

Galindo-Leon, E. E., Lin, F. G., and Liu, R. C. (2009). Inhibitory plasticity in a lateral band improves cortical detection of natural vocalizations. *Neuron* 62, 705–716. doi: 10.1016/j.neuron.2009.05.001

Ghose, G. M., and Maunsell, J. H. R. (2002). Attentional modulation in visual cortex depends on task timing. *Nature* 419, 616–620. doi: 10.1038/nature01057

Gomez, P., Ratcliff, R., and Perea, M. (2007). A Model of the go/no-go task. *J. Exp. Psychol. Gen.* 136, 389–413.

Grant, K. W., and Seitz, P.-F. (2000). The use of visible speech cues for improving auditory detection of spoken sentences. *J. Acoust. Soc. Am.* 108, 1197–1208. doi: 10.1121/1.1288668

Green, D. M., and Swets, J. A. (1966). *Signal Detection Theory and Psycho-Physics*. New York, NY: Wiley.

Greenberg, G. Z., and Larkin, W. D. (1968). Frequency-response characteristic of auditory observers detecting signals of a single frequency in noise: the probe-signal method. *J. Acoust. Soc. Am.* 44, 1513–1523. doi:10.1121/1.1911290

Guo, W., Clause, A. R., Barth-maron, A., Polley, D. B., Guo, W., Clause, A. R., et al. (2017). A corticothalamic circuit for dynamic switching between feature detection and discrimination. *Neuron* 95, 180–194. doi: 10.1016/j.neuron.2017.05.019

Guo, Z. V., Hires, S. A., Li, N., O'Connor, D. H., Komiyama, T., Ophir, E., et al. (2014). Procedures for behavioral experiments in head-fixed mice. *PLoS One* 9:e88678. doi: 10.1371/journal.pone.0088678

Gutschalk, A., Micheyl, C., and Oxenham, A. J. (2008). Neural correlates of auditory perceptual awareness under informational masking. *PLoS Biol.* 6:1156. doi: 10.1371/journal.pbio.0060138

Homma, N. Y., Happel, M. F. K., Nodal, F. R., Ohl, F. W., King, A. J., and Bajo, V. M. (2017). A role for auditory corticothalamic feedback in the perception of complex sounds. *J. Neurosci.* 37, 6149–6161. doi: 10.1523/jneurosci.0397-17.2017

Janssen, P., and Shadlen, M. N. (2005). A representation of the hazard rate of elapsed time in macaque area LIP. *Nat. Neurosci.* 8, 234–241. doi: 10.1038/nn1386

Jaramillo, S., and Zador, A. M. (2011). The auditory cortex mediates the perceptual effects of acoustic temporal expectation. *Nat. Neurosci.* 14, 246–251. doi: 10.1038/nn.2688

Joris, P. X., Schreiner, C. E., and Rees, A. (2004). Neural processing of amplitude-modulated sounds. *Physiol. Rev.* 84, 541–577. doi: 10.1152/physrev.00029.2003

Kidd, G., Mason, C. R., Deliwala, P. S., Woods, W. S., and Colburn, H. S. (1994). Reducing informational masking by sound segregation. *J. Acoust. Soc. Am.* 95, 3475–3480. doi: 10.1121/1.410023

Kuchibhotla, K. V., Gill, J. V., Lindsay, G. W., Papadoyannis, E. S., Field, R. E., Sten, T. A. H., et al. (2017). Parallel processing by cortical inhibition enables context-dependent behavior. *Nat. Neurosci.* 20, 62–71. doi: 10.1038/nn.4436

Lakatos, P., Karmos, G., Mehta, A. D., Ulbert, I., and Schroeder, C. E. (2008). Entrainment of neuronal oscillations as a mechanism of attentional selection. *Science* 320, 110–113. doi: 10.1126/science.1154735

Lakatos, P., Musacchia, G., O'Connel, M. N., Falchier, A. Y., Javitt, D. C., and Schroeder, C. E. (2013). The spectrotemporal filter mechanism of auditory selective attention. *Neuron* 77, 750–761. doi: 10.1016/j.neuron.2012.11.034

Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *J. Acoust. Soc. Am.* 49, 467–477. doi: 10.1121/1.1912375

MacDonald, C. J., Carrow, S., Place, R., and Eichenbaum, H. (2013). Distinct hippocampal time cell sequences represent odor memories in immobilized Rats. *J. Neurosci.* 33, 14607–14616. doi: 10.1523/jneurosci.1537-13.2013

McDermott, J. H., Wrobleski, D., and Oxenham, A. J. (2011). Recovering sound sources from embedded repetition. *PNAS* 108, 1188–1193. doi: 10.1073/pnas.1004765108

McGinley, M. J., David, S. V., and McCormick, D. A. (2015). Cortical membrane potential signature of optimal states for sensory signal detection. *Neuron* 87, 179–192. doi: 10.1016/j.neuron.2015.05.038

Mesgarani, N., and Chang, E. F. (2012). Selective cortical representation of attended speaker in multi-talker speech perception. *Nature* 485, 233–236. doi: 10.1038/nature11020

Micheyl, C., Shamma, S. A., and Oxenham, A. J. (2007). "Hearing out repeating elements in randomly varying multitone sequences: a case of streaming?," in *Hearing – From Sensory Processing to Perception*, eds B. Kollmeier, G. Klump, V. Hohmann, U. Langemann, M. Mauermann, S. Uppenkamp, et al. (Berlin: Springer).

Morillon, B., Hackett, T. A., Kajikawa, Y., and Schroeder, C. E. (2015). Predictive motor control of sensory dynamics in auditory active sensing. *Curr. Opin. Neurobiol.* 31C, 230–238. doi: 10.1016/j.conb.2014.12.005

Nakajima, M., Schmitt, L. I., and Halassa, M. M. (2019). Prefrontal cortex regulates sensory filtering through a basal ganglia-to-thalamus pathway. *Neuron* 103, 445.e10–458.e10.

Narain, D., Remington, E. D., Zeeuw, C. I. D., and Jazayeri, M. (2018). A cerebellar mechanism for learning prior distributions of time intervals. *Nat. Commun.* 9, 1–12.

Nobre, A. C., and Van Ede, F. (2018). Anticipated moments: temporal structure in attention. *Nat. Rev. Neurosci.* 19, 34–48. doi: 10.1038/nrn.2017.141

O'Connell, M. N., Falchier, A., McGinnis, T., Schroeder, C. E., and Lakatos, P. (2011). Dual mechanism of neuronal ensemble inhibition in primary auditory cortex. *Neuron* 69, 805–817. doi: 10.1016/j.neuron.2011.01.012

Piéron, H. (1913). II. Recherches sur les lois de variation des temps de latence sensorielle en fonction des intensités excitatrices. *Ann. Psychol.* 20, 17–96. doi: 10.3406/psy.1913.4294

Reznik, D., Henkin, Y., Schadel, N., and Mukamel, R. (2014). Lateralized enhancement of auditory cortex activity and increased sensitivity to self-generated sounds. *Nat. Commun.* 5:5049.

Reznik, D., Simon, S., and Mukamel, R. (2018). Predicted sensory consequences of voluntary actions modulate amplitude of preceding readiness potentials. *Neuropsychologia* 119, 302–307. doi: 10.1016/j.neuropsychologia.2018.08.028

Rodgers, C. C., and DeWeese, M. R. (2014). Neural correlates of task switching in prefrontal cortex and primary auditory cortex in a novel stimulus selection task for rodents. *Neuron* 82, 1157–1170. doi: 10.1016/j.neuron.2014.04.031

Romero, S., Hight, A. E., Clayton, K. K., Resnik, J., Williamson, R. S., Hancock, K. E., et al. (2020). Cellular and widefield imaging of sound frequency organization in primary and higher order fields of the mouse auditory cortex. *Cereb. Cortex* 30, 1603–1622. doi: 10.1093/cercor/bhz190

Rummell, B. P., Klee, J. L., and Sigurdsson, T. (2016). Attenuation of responses to self-generated sounds in auditory cortical neurons. *J. Neurosci.* 36, 12010–12026. doi: 10.1523/jneurosci.1564-16.2016

Runyan, C. A., Piasini, E., Panzeri, S., and Harvey, C. D. (2017). Distinct timescales of population coding across cortex. *Nature* 548, 92–96. doi: 10.1038/nature23020

Sanders, J. I., and Kepecs, A. (2012). Choice ball: a response interface for two-choice psychometric discrimination in head-fixed mice. *J. Neurophysiol.* 108, 3416–3423. doi: 10.1152/jn.00669.2012

Schneider, D. M., Sundararajan, J., and Mooney, R. (2018). A cortical filter that learns to suppress the acoustic consequences of movement. *Nature* 561, 391–395. doi: 10.1038/s41586-018-0520-5

Shepard, K. N., Chong, K. K., and Liu, R. C. (2016). Contrast enhancement without transient map expansion for species-specific vocalizations in core auditory cortex during learning. *eNeuro* 3:ENEURO.0318-16.2016. doi: 10.1523/ENEURO.0318-16.2016

Siemann, J. K., Muller, C. L., Bamberger, G., Allison, J. D., Veenstra-VanderWeele, J., and Wallace, M. T. (2015). A novel behavioral paradigm to assess multisensory processing in mice. *Front. Behav. Neurosci.* 8:456. doi: 10.3389/fnbeh.2014.00456

Sohoglu, E., and Chait, M. (2016). Detecting and representing predictable structure during auditory scene analysis. *eLife* 5, 1–17.

Southwell, R., Baumann, A., Gal, C., Barascud, N., Friston, K., and Chait, M. (2017). Is predictability salient? A study of attentional capture by auditory patterns. *Philos. Trans. R. Soc.* 372:20160105. doi: 10.1098/rstb.2016.0105

Stanislaw, H., and Todorov, N. (1999). Calculation of signal detection theory measures. *Behav. Res. Methods Instruments Comput.* 3, 137–149. doi: 10.3758/bf03207704

Stebbings, K. A., Lesicko, A. M. H., and Llano, D. A. (2014). The auditory corticocollicular system: molecular and circuit-level considerations. *Hear. Res.* 314, 51–59. doi: 10.1016/j.heares.2014.05.004

Stevenson, R. A., Zemtsov, R. K., and Wallace, M. T. (2011). Multisensory illusions and the temporal binding window. *iPerception* 2, 903–903. doi: 10.1068/ic903

Suga, N., and Ma, X. (2003). Multiparametric corticofugal modulation and plasticity in the auditory system. *Nat. Rev. Neurosci.* 4, 783–794. doi: 10.1038/nrn1222

Sumby, W. H., and Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *J. Acoust. Soc. Am.* 26, 212–215. doi: 10.1121/1.1907309

Takahashi, N., Oertner, T. G., Hegemann, P., and Larkum, M. E. (2016). Active cortical dendrites modualte perception. *Science* 354, 1159–1165.

Tsunada, J., Liu, A. S. K., Gold, J. I., and Cohen, Y. E. (2016). Causal contribution of primate auditory cortex to auditory perceptual decision-making. *Nat. Neurosci.* 19, 135–142. doi: 10.1038/nn.4195

van Wassenhove, V., Grant, K. W., and Poeppel, D. (2007). Temporal window of integration in auditory-visual speech perception. *Neuropsychologia* 45, 598–607. doi: 10.1016/j.neuropsychologia.2006.01.001

Viemeister, N. F., and Wakefield, G. H. (1991). Temporal integration and multiple looks. *J. Acoust. Soc. Am.* 90, 858–865. doi: 10.1121/1.401953

Vincis, R., Chen, K., Czarnecki, L., Chen, J., and Fontanini, A. (2020). Dynamic representation of taste-related decisions in the gustatory insular cortex of mice. *Curr. Biol.* 30, 1834.e5–1844.e5.

Wiegand, K., and Gutschalk, A. (2012). Correlates of perceptual awareness in human primary auditory cortex revealed by an informational masking experiment. *Neuroimage* 61, 62–69. doi: 10.1016/j.neuroimage.2012.02.067

Williamson, R. S., and Polley, D. B. (2019). Parallel pathways for sound processing and functional connectivity among layer 5 and 6 auditory corticofugal neurons. *eLife* 8:e42974.

Winer, J. A. (2005). Decoding the auditory corticofugal systems. *Hear. Res.* 207, 1–9. doi: 10.1016/j.heares.2005.06.007

Wright, B. A., and Fitzgerald, M. B. (2004). The time course of attention in a simple auditory detection task. *Percept. Psychophys.* 66, 508–516. doi: 10.3758/bf03194897

Xiong, X. R., Liang, F., Zingg, B., Ji, X. Y., Ibrahim, L. A., Tao, H. W., et al. (2015). Auditory cortex controls sound-driven innate defense behaviour through corticofugal projections to inferior colliculus. *Nat. Commun.* 6, 1–12. doi: 10.1016/0378-5955(88)90047-0

Xu, N. L., Harnett, M. T., Williams, S. R., Huber, D., O'Connor, D. H., Svoboda, K., et al. (2012). Nonlinear dendritic integration of sensory and motor input during an active sensing task. *Nature* 492, 247–251. doi:10.1038/nature11601

Yilmaz, M., and Meister, M. (2013). Rapid innate defensive responses of mice to looming visual stimuli. *Curr. Biol.* 23, 2011–2015. doi: 10.1016/j.cub.2013.08.015

Yin, P., Elhilali, M., Fritz, J. B., and Shamma, S. (2007). "Primary auditory cortical responses while attending to different streams," in *Hearing – From Sensory Processing to Perception*, eds B. Kollmeier, G. Klump, V. Hohmann, U. Langemann, M. Mauermann, S. Uppenkamp, et al. (Berlin: Springer).

Zhang, S., Xu, M., Kamigaki, T., Do, J. P. H., Chang, W.-C., Jenvay, S., et al. (2014). Long-range and local circuits for top-down modulation of visual cortex processing. *Science* 345, 660–665. doi: 10.1126/science.1254126

Zhou, S., Masmanidis, S. C., and Buonomano, D. V. (2020). Neural sequences as an optimal dynamical regime for the readout of time. *Neuron* 108, 651.e5–658.e5.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Clayton, Asokan, Watanabe, Hancock and Polley. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 22 July 2021
doi: 10.3389/fnins.2021.693935

Excitatory Repetitive Transcranial Magnetic Stimulation Over Prefrontal Cortex in a Guinea Pig Model Ameliorates Tinnitus

*Jack W. Zimdahl[1], Harrison Thomas[1], Samuel J. Bolland[2,3], Kerry Leggett[1], Kristin M. Barry[1], Jennifer Rodger[2,3] and Wilhelmina H. A. M. Mulders[1]**

[1] School of Human Sciences, University of Western Australia, Crawley, WA, Australia, [2] School of Biological Sciences, University of Western Australia, Crawley, WA, Australia, [3] Perron Institute for Neurological and Translational Research, Crawley, WA, Australia

OPEN ACCESS

Edited by:
David K. Ryugo, Garvan Institute of Medical Research, Australia

Reviewed by:
Jinsheng Zhang, Wayne State University, United States
Marlies Knipper, University of Tübingen, Germany

***Correspondence:**
Wilhelmina H. A. M. Mulders
helmy.mulders@uwa.edu.au

Specialty section:
This article was submitted to Auditory Cognitive Neuroscience, a section of the journal Frontiers in Neuroscience

Received: *12 April 2021*
Accepted: *01 July 2021*
Published: *22 July 2021*

Citation:
Zimdahl JW, Thomas H, Bolland SJ, Leggett K, Barry KM, Rodger J and Mulders WHAM (2021) Excitatory Repetitive Transcranial Magnetic Stimulation Over Prefrontal Cortex in a Guinea Pig Model Ameliorates Tinnitus. Front. Neurosci. 15:693935. doi: 10.3389/fnins.2021.693935

Tinnitus, a phantom auditory perception that can seriously affect quality of life, is generally triggered by cochlear trauma and associated with aberrant activity throughout the auditory pathways, often referred to as hyperactivity. Studies suggest that non-auditory structures, such as prefrontal cortex (PFC), may be involved in tinnitus generation, by affecting sensory gating in auditory thalamus, allowing hyperactivity to reach the cortex and lead to perception. Indeed, human studies have shown that repetitive transcranial magnetic stimulation (rTMS) of PFC can alleviate tinnitus. The current study investigated whether this therapeutic effect is achieved through inhibition of thalamic hyperactivity, comparing effects of two common clinical rTMS protocols with sham treatment, in a guinea pig tinnitus model. Animals underwent acoustic trauma and once tinnitus developed were treated with either intermittent theta burst stimulation (iTBS), 20 Hz rTMS, or sham rTMS (10 days, 10 min/day; weekdays only). Tinnitus was reassessed and extracellular recordings of spontaneous tonic and burst firing rates in auditory thalamus made. To verify effects in PFC, densities of neurons positive for calcium-binding proteins, calbindin and parvalbumin, were investigated using immunohistochemistry. Both rTMS protocols significantly reduced tinnitus compared to sham. However, spontaneous tonic firing decreased following 20 Hz stimulation and increased following iTBS in auditory thalamus. Burst rate was significantly different between 20 Hz and iTBS stimulation, and burst duration was increased only after 20 Hz treatment. Density of calbindin, but not parvalbumin positive neurons, was significantly increased in the most dorsal region of PFC indicating that rTMS directly affected PFC. Our results support the involvement of PFC in tinnitus modulation, and the therapeutic benefit of rTMS on PFC in treating tinnitus, but indicate this is not achieved solely by suppression of thalamic hyperactivity.

Keywords: repetitive transcranial magnetic stimulation, tinnitus, guinea pig, prefrontal cortex, hyperactivity, calcium-binding protein, medial geniculate nucleus

INTRODUCTION

Tinnitus is a phantom auditory percept, often described as a ringing or buzzing in the ear, which affects 10–15% of the population (Baguley et al., 2013). For about 1.5% of sufferers tinnitus severely affects quality of life, leading to loss of sleep and concentration, anxiety, stress, and suicidal ideation (Baguley et al., 2013). The precise neural substrate is still debated (Baguley et al., 2013), and different mechanisms may exist for subtypes of tinnitus (Vanneste et al., 2019), both factors contributing to the lack of a current universal cure. Tinnitus is commonly associated with cochlear damage and some degree of hearing loss. This leads to changes in central auditory structures (Tan et al., 2013), such as increased neural synchrony, and increased spontaneous and bursting activity (Norena and Eggermont, 2003; Mulders and Robertson, 2009; Kalappa et al., 2014; Basura et al., 2015; Wu et al., 2016), which may all contribute to tinnitus generation (Eggermont and Roberts, 2015).

Alongside its complex ascending projections, the auditory system also encompasses an elaborate system of descending pathways (Schofield, 2011). These descending pathways allow for modulation of the auditory information *en route* to cortex and are therefore in a prime position to affect auditory perception (see for example: Guo et al., 2017; Homma et al., 2017; GuinanJr., 2018). In addition, efferent modulation of ascending auditory information also arises from non-auditory brain structures involved in emotional and memory processing such as the prefrontal cortex (PFC), amygdala and hippocampus (Scannell et al., 1999; Rolls, 2015), further contributing to the conscious perception of auditory information (McCormick, 1992). More specifically, it is proposed that these non-auditory inputs play an important role in sensory gating, inhibiting non-salient signals (Rauschecker et al., 2010; De Ridder et al., 2014). This has now led to the hypothesis that failure of gating for non-salient signals may be involved in the generation of tinnitus (Rauschecker et al., 2010, 2015; De Ridder et al., 2014).

Emerging data suggest that tinnitus may be due to the breakdown of sensory gating circuitry between the auditory system and non-auditory limbic regions (Rauschecker et al., 2010; De Ridder et al., 2014), such that altered activity due to cochlear damage reaches the cortex and ultimately leads to a conscious percept (Rauschecker et al., 2010, 2015; De Ridder et al., 2014). Specifically it has been suggested when the increased spontaneous activity in tinnitus patients fails to be ignored and is recognized as sound, this leads to a resetting of auditory predictions allowing the tinnitus to continue (Sedley et al., 2016, 2019). Debate does exist to the cause of the increased spontaneous activity as some suggest an increased neural gain (Sheppard et al., 2020) whereas others argue the increased neural gain is underlying hyperacusis but not tinnitus (Brotherton et al., 2015).

Abbreviation: AT, acoustic trauma; CAP, compound action potential; ECG, electrocardiogram; E-fields, electric fields; GPIAS, gap prepulse inhibition of the acoustic startle; iTBS, intermittent theta burst stimulation; IQR, interquartile range; MGN, medial geniculate nucleus; PFC, prefrontal cortex; PPI, prepulse inhibition; ROI, region of interest; rTMS, repetitive transcranial magnetic stimulation; RMS, root mean square.

Part of the proposed circuitry involved in sensory gating involves indirect pathways from ventromedial PFC and/or anterior cingulate cortex to auditory thalamus (medial geniculate nucleus [MGN]) (O'Donnell et al., 1997; Anderson et al., 2006; Zikopoulos and Barbas, 2006; Yu et al., 2009; Crippa et al., 2010; Smith et al., 2012; Bartlett, 2013; Mellott et al., 2014; KeiferJr., Gutman et al., 2015; Rauschecker et al., 2015). Indeed, we have shown in rats that neural activity in MGN can be modulated by input from nucleus accumbens and PFC (Barry et al., 2015, 2017). Rodent PFC as targeted in our studies is thought to be analogous to ventromedial PFC and anterior cingulate in humans (Bicks et al., 2015; Laubach et al., 2018).

Clinical studies in tinnitus patients have shown that non-invasive stimulation of PFC, using techniques such as repetitive transcranial stimulation (rTMS) and direct current stimulation, can alleviate tinnitus loudness and distress (Kleinjung et al., 2008; Vanneste et al., 2011; De Ridder et al., 2013; Lehner et al., 2013; Langguth et al., 2014). Indeed this was recently replicated in a randomised placebo-controlled, single-blinded clinical trial using high frequency rTMS over the dorsomedial PFC, which showed a statistically significant reduction in tinnitus severity (Ciminelli et al., 2020). However, these human studies cannot establish whether these beneficial effects are due to inhibition of thalamic hyperactivity as suggested by the dysfunctional gating theory (Rauschecker et al., 2010; De Ridder et al., 2014). In the present study, we implemented our guinea pig model of acoustic trauma (AT) and tinnitus (Mulders et al., 2014a, 2016) to investigate the effects of two excitatory rTMS protocols, 20 Hz and intermittent theta burst stimulation (iTBS) applied over PFC, and assessed neuronal firing patterns in MGN as well as behavioral signs of tinnitus. The MGN was selected as a target to measure neuronal activity as we have previously shown that it shows increased spontaneous firing rates in the presence of tinnitus (Cook et al., 2021) and electrical stimulation of PFC evokes changes in MGN neuronal firing patterns (Barry et al., 2017). Modeling of the rTMS coil was used to assess the extent and strength of the electric fields (E-fields) and to investigate whether the MGN could be directly activated. Excitation of PFC was hypothesized to have inhibitory effects on thalamic and therefore cortical activity, and in this way, attenuate tinnitus. In addition, we investigated the effects of rTMS on PFC, through quantitative immunohistochemical analysis of calcium binding proteins, as studies have shown changes in these proteins following rTMS treatment (Benali et al., 2011; Hoppenrath and Funke, 2013; Volz et al., 2013; Labedi et al., 2014; Mix et al., 2014; Mulders et al., 2019).

EXPERIMENTAL PROCEDURES

Animals and Experimental Design

All experiments were in line with the Code of the National Health and Medical Research Council of Australia and the National Institutes of Health Guide for the Care and Use of Laboratory Animals (NIH Publications No. 80-23, revised 1996) with approval from the Animal Ethics Committee of The University of Western Australia (RA/3/100/1458). All efforts

were made to minimise the number of animals used and their suffering. Twenty-seven pigmented adult guinea pigs of either sex (20 males; 7 females) were used. One additional guinea pig (female 875 g) was used for MRI imaging.

Detailed descriptions of behavioral testing for tinnitus (Robertson et al., 2013; Mulders et al., 2014a, 2016; Leggett et al., 2018); AT surgery (Mulders and Robertson, 2009; Mulders et al., 2011, 2014a,b; Mulders and Robertson, 2011; Vogler et al., 2011; Robertson et al., 2013); electrophysiological recordings in MGN including burst firing analysis (Barry et al., 2019; Mulders et al., 2019); and immunohistochemistry with associated quantitative analysis to determine densities of calcium-binding neurons in PFC (Mulders et al., 2019) have been described in previous papers from our laboratory and will be described only in brief in the next sections.

Behavioral Tests for Tinnitus

Behavioral testing for tinnitus consisted of gap prepulse inhibition of acoustic startle (GPIAS) in combination with prepulse inhibition (PPI). Tinnitus testing was performed before (to obtain baseline measures) and after AT for all animals. There was at least 1 day between testing sessions and a maximum of three testing days per week to limit habituation effects. During behavioral testing, animals were lightly restrained in clear polycarbonate holders and placed on force transducing platforms in a soundproof room. Animals were left to acclimatize for 5 min prior to each testing session. Each animal was allocated to the same platform for all subsequent testing.

Both GPIAS and PPI involve the delivery of a startle tone after which the amplitude of the animals' startle response is recorded. The startle stimulus was identical in both PPI and GPIAS conditions (1 kHz, 0.5k Hz bandwidth; 106 dB SPL; 50 ms duration) and was delivered by a speaker positioned approximately 5 cm above the animals' head (Radio Shack 401278B). An additional speaker (Beyer DT 48) was positioned approximately 3 cm above the animals' head to deliver the continuous background noise (GPIAS), or prepulse stimulus (PPI).

Prepulse inhibition occurs when a weak prepulse stimulus inhibits the startle response to a succeeding stronger stimulus. The PPI sessions consisted of 50 trials. In half of the trials, a 50 ms prepulse was presented 100 ms before the delivery of the startle tone, in the other half, no prepulse was present. The prepulse was a narrowband noise centered at either 8 kHz (10 dB bandwidth 2.2 kHz) or 14 kHz (10 dB bandwidth 1.6 kHz) at 66 dB SPL intensity. These two center frequencies were selected as they are either below the AT frequency in a compound action potential (CAP) audiogram region not showing threshold loss or just above the AT frequency in a CAP audiogram region showing significant threshold loss. The interval between startle stimulus presentations varied randomly by 20–30 s, and the order of the prepulse and non-prepulse trials was randomized.

Gap prepulse inhibition of the acoustic startle testing is a variant of PPI in which a silent gap, functioning as a prepulse, is inserted in a continuous background noise preceding the startle tone. GPIAS also consisted of 50 trials, with half of the trials incorporating a gap. The background noise was at the same level and characteristics as the pre-pulses in the PPI test. During testing, the animals' startle is measured from a startle platform output and is calculated as root mean square (RMS) of force produced during the baseline response (before startle) and startle response.

For analysis, the startle response of prepulse trials was compared with no prepulse trials (for PPI) or gap trials with no gap trials (for GPIAS) within each animal. Outliers in trials were identified through RMS values $\pm$ 3 SD from the mean. Mean PPI or GPIAS suppression was expressed in percentage by comparing the RMS force between the prepulse and no prepulse trials or between the gap and no gap trials, respectively. An animal is considered to "pass" if there is a significant difference (Mann–Whitney test, $p < 0.05$) between prepulse/no prepulse or gap/no gap trials, and "fail" if the condition is not met (Mann–Whitney test, $p > 0.05$). All animals passed the PPI test once and the GPIAS paradigm twice indicating stable baselines before AT.

Possible tinnitus development was assessed after AT surgery via weekly GPIAS testing. Animals may fail GPIAS due to tinnitus but also alternatively because of either hearing loss or deficits in the neural circuitry underlying startle response and PPI and hence animals were required to pass PPI after AT to ensure that failure of GPIAS was specifically related to tinnitus. Animals that failed GPIAS on two repeat occasions (at least 1 day between sessions) and passed PPI, were considered to have behavioral signs of tinnitus. For group comparisons, GPIAS suppression was averaged over two sessions before the AT, at the time of tinnitus development and after sham or active TMS treatment preceding the final electrophysiological experiment.

Acoustic Trauma Surgery

Acoustic trauma surgery consisted of opening the tympanic bulla to enable placement of a recording electrode on the round window of the cochlea and the measurement of thresholds of the auditory nerve before and after an AT. For anesthesia, animals received a subcutaneous injection of atropine 0.1 ml (Atropine, atropine sulfate 0.6 mg/ml, Apex Laboratories, Somersby, Australia) followed by 5 mg/kg Diazepam (Pamlin, diazepam 5 mg/ml, Ceva Animal Health, Glenorie, Australia) intraperitoneally, an intramuscular injection of 1 ml/kg Hypnorm (Hypnorm, 0.315 mg/ml fentanyl citrate and 10 mg/ml fluanisone, VetaPharma, Leeds, United Kingdom), and a subcutaneous injection of 0.1 ml lignocaine (Lignocaine20, 20 mg/ml lignocaine HCl, Troy Laboratories, Glendenning, Australia). Once surgical anesthesia was obtained, animals were placed on a heated platform in a soundproof room and mounted in hollow ear bars. Anesthesia level was maintained throughout surgery with an additional administration (one-third of the initial dose) of Hypnorm.

To assess hearing, a small incision and small hole in the tympanic bulla allowed for an insulated silver wire to be placed onto the round window. CAP thresholds were measured in response to pure tone stimuli (10 ms duration, 4/s, frequency range: 4–24 kHz) created in a closed sound system using a 1/2 inch condenser microphone driven in reverse (Bruel and Kjaer, type 4134). CAP signals were recorded using a custom-made computer program (sample rate 96 kHz, Neurosound; MI Lloyd).

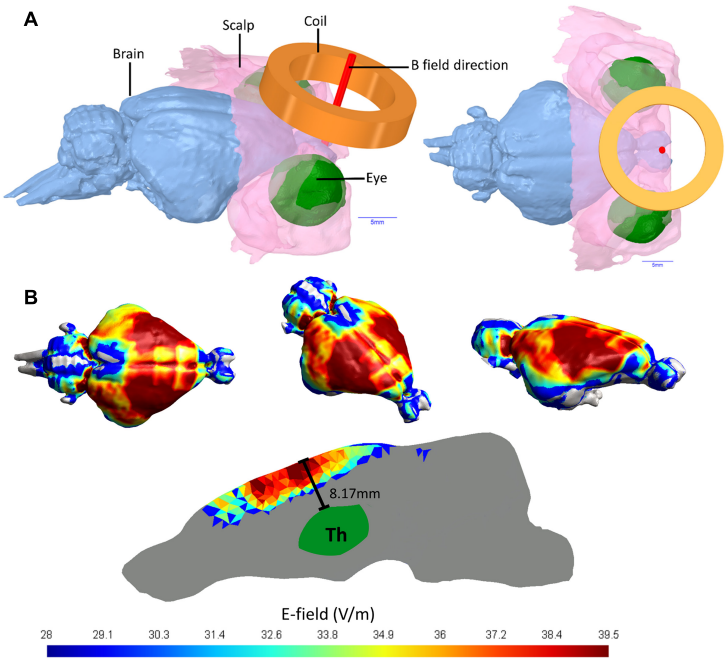

FIGURE 1 | Guinea pig head 3D model and E-fields in guinea pig brain. **(A)** MRI imaging of a Guinea pig head utilized to produce a 3D head model showing the brain, the coil position held 2 mm from the scalp and the E-field direction of the active coil pointing directly into the brain. **(B)** The E-fields in V/m for simulated TMS stimulation of a guinea pig brain using a MagVenture Cool-40 Rat coil at 23% of maximum stimulator output.

CAP signals were amplified (1000x), filtered (100 Hz – 3 kHz bandpass) and recorded (Powerlab 4SP, AD instruments). Then a unilateral AT (left ear, 10 kHz, 124 dB SPL, 2 h) was performed, whilst the contralateral ear (right ear) was blocked with plasticine. After AT, another CAP audiogram was measured, the incision was sutured, and animals were allowed to recover.

rTMS Protocols

rTMS was delivered using a commercially available animal-specific coil (Cool-40 Rat Coil; Magventure, Farum, Denmark). Machine stimulus output was set to 23% to avoid direct facial muscle twitching (assessed visually). Treatment was performed on awake animals whilst being gently held in the experimenter's lap for 10 min daily, on weekdays only, over a 2-week period. The peak induced electric field was marked allowing for positioning of the coil to be placed against the animals head over the PFC 14.2 mm anterior of the interaural line (Rapisarda and Bacchelli, 1977). This was achieved through shaving the animals' head and marking the position with a permanent marker using handheld electronic calipers measuring from the interaural line; the interaural line was judged using stretched tape between the center of the ear canals.

Two excitatory stimulatory rTMS protocols that were based on clinical settings (Huang et al., 2005; Kleinjung et al., 2008; Lehner et al., 2013; Langguth et al., 2014) and that were similar in duration and pulse number, were used. The first, 20 Hz (2000 pulses in 10 min stimulation) was applied in 40 blocks of 50 pulses with 13 s intervals to avoid coil overheating. The other protocol, iTBS, included 2 s of stimulation trains, 10 pulses of three bursts each at 50 Hz, with an 8 s inter-train interval (1800 pulses over a 10 min period). Sham stimulation was delivered at 20 Hz, but with the coil positioned 20 cm above the animals' head, and angled perpendicular to the animals' longitudinal axis. The sham rTMS procedure generated the same sound as verum rTMS, but gauss meter recordings at the same position as the animals' heads produced no detectable magnetic field. At the same time a cardboard sham coil with similar dimensions and color was held on the head to simulate coil placement sensation.

rTMS E-Field Modelling

The rTMS based E-fields induced in guinea pig brains in this current study were simulated using 3D computer models to investigate the E-field spread and intensity. The 3D head model was produced from anatomical T2 weighted MRI images with 100 μm thick isotropic voxels (336 slices) of a guinea pig head, which were attained using a 9.4T Bruker Biospec 94/30 small animal MRI machine. The head model was segmented into different tissue types to account for differences in conductivity. The brain was segmented out using FMRIB software library (Jenkinson et al., 2012) via the function Brain Extraction Tool (BET), while the non-brain tissues were segmented out using the program ITK-SNAP (Yushkevich et al., 2016) based on MRI image intensity values. ITK-SNAP was used to produce 3D surfaces and the software package ANSYS Academic (SpaceClaim 2019 R3, Release 19.5.0) was used to build the 3D mesh with further optimization through MeshFix (Attene, 2010). The triangle and tetrahedral based 3D FEM brain model was produced in Gmsh and the resulting model contains ∼7,152,380

tetrahedral elements and ∼1,290,940 triangles. The whole brain volume is 2.76 cm^3 and Coil distance was set at 2 mm above the scalp to account for fur (**Figure 1A**). Coil definition files were provided by the program SimNIBS 3.1.2 (Saturnino et al., 2019), which was also used to solve the finite element method based rTMS simulation. The Cool-40 Rat coil file analysis, model processing and SimNIBS output analysis were conducted in MATLAB (R2016a & R2017a, The Mathworks, Inc., Natick, MA, United States). SimNIBS uses GetDP solver (Geuzaine, 2007) to calculate simulated rTMS induced E-field values. Default isotropic conductivities were set for the soft tissues (σ = 0.465 S/m), skull (σ = 0.01 S/m), CSF (σ = 1.654 S/m), gray matter (σ = 0.275 S/m), and white matter (σ = 0.126 S/m) (Wagner et al., 2004). E-field distributions were calculated with the Cool-40 Rat coil positioned tangentially to scalp surface and the stimulator output was set to 23% of maximum stimulator output as per the rTMS protocols. E-field spread over the brain tissue would be the same for the iTBS and 20 Hz protocol.

Electrophysiological Recordings and Burst Firing Analysis

Three days after cessation of the sham or active rTMS treatments, a final GPIAS test was performed and followed by an electrophysiological experiment using extracellular recordings to assess MGN neuronal activity. During this non-recovery surgery animals received a 0.1 ml subcutaneous injection of atropine and an intraperitoneal injection of 30 mg/kg pentobarbitone sodium (Pentobarbitone, pentobarbitone sodium 60 mg/ml, Troy Laboratories, Glendenning, Australia), followed by a 0.15 ml (fixed dose) intramuscular injection of Hypnorm. To maintain anesthesia 0.15 ml of Hypnorm was given hourly and half doses of pentobarbitone sodium were given every 2 h. A tracheotomy was performed, and animals artificially ventilated on carbogen (95% O_2, 5% CO_2). Animals were placed on a heated platform in a soundproof room and mounted in hollow ear bars. An electrocardiogram (ECG) was recorded throughout surgery to ensure depth of anesthesia and monitor physiological status. ECG was expressed by measurements of the interval between QRS complexes (ECG interval) which were continuously displayed during the experiment on an oscilloscope. In a previous experiment (Cook et al., 2021) we have demonstrated that spontaneous firing rates in MGN significantly decrease with ECG intervals > 300 ms. Therefore, recordings were only included and analyzed while the ECG intervals were under 300 ms. At the conclusion of the experiment, animals were euthanized via an injection of fixed dose of 0.3 ml Lethabarb (sodium pentobarbitone 325 mg/ml; Virbac, Milperra, Australia).

Additionally, during this non-recovery surgery bilateral CAP audiograms were measured using the method detailed above (AT surgery). Following CAP audiograms, a partial craniotomy was performed over the cortex overlying the right MGN, contralateral to the AT exposed ear (Rapisarda and Bacchelli, 1977). The craniotomy was covered with 5% agar in saline to improve stability of recordings and prevent dehydration of the neural tissue. The contralateral right ear was blocked with plasticine. A glass insulated tungsten microelectrode was advanced along the dorso-ventral axis through the cortex to the MGN. Entry into MGN was indicated by noise-evoked cluster activity, approximately 6–7 mm from the cortical surface. Single neuron characteristic frequency (CF) and threshold were determined audio-visually using the Neurosound software when possible, and the spontaneous firing rate of the neurons was measured during a 10 s period. Immediately after these measurements, 2 min of spontaneous firing rate were recorded on LabChart (ADInstruments) in the absence of sound, to allow for post-experimental analysis of burst firing. Single neuron recordings continued until at least 50 neurons were recorded per animal or until the ECG intervals exceeded 300 ms.

Recording location in MGN was confirmed via histological analysis of electrode tracks (data not shown). In each animal 4–8 tracks were required to collect sufficient recordings. Onset latencies varied considerably from 12 to >200 ms. Neuronal CF could not always be determined as some neurons responded only to broadband stimuli. In addition, a very small number of neurons could not be driven by sound, but since they were found in between other neurons that did show a robust response to sound, they were still included in analysis. No attempt was made to identify the subdivision of the MGN. Subdivision identification would have required more recording time per neuron, and this was not compatible with our goal to collect data from 50 neurons per animal whilst maintaining the required physiological status.

Burst Firing Analysis

Individual neurons were isolated based on amplitude and wave shape from the 2-min neuronal recordings, using the LabChart spike histogram software (ADInstruments). The individual neuron data were imported into NeuroExplorer v.4.135 (2014) software. Analysis of burst firing was performed using the burst analysis function which measured bursts per minute, and mean burst duration (Kepecs and Lisman, 2004; Kalappa et al., 2014; Kimura and Imbe, 2015). Burst firing criteria were selected based on previous studies and on visual inspection of the bursting patterns observed on the LabChart (ADInstruments) files and were set as follows: maximum interval to start burst (8 ms), maximum interval to end burst (8 ms), minimum interval between bursts (15 ms), minimum duration of burst (6 ms), and minimum number of spikes per burst (3). Neurons were only included if they were clearly distinguishable from other neurons in the recording and spike height was constant throughout the 2 min of recording. In addition, only neurons with a spontaneous firing rate > 1 spike/sec were used for subsequent burst firing analysis.

Histology and Immunohistochemistry

Following electrophysiological recordings, animals were euthanized (0.4 ml intraperitoneal injection of Lethabarb – sodium pentobarbitone 325 mg/ml; Virbac, Milperra, Australia), and transcardially perfused with saline followed by 4% paraformaldehyde in 0.1 M phosphate buffer. Brains were removed and stored in fixative overnight followed by 30% sucrose in 0.1 M phosphate buffer solution for 48 h at 4°C. Sections were cut at 60 μm using a freezing microtome, mounted on gelatine coated slides, stained with Toluidine

Blue and coverslipped. Electrode tracts were photographed using light microscopy (Nikon Eclipse 80i) connected to a camera (DigiSight) using NIS Elements Advanced Research software (Nikon).

Immunohistochemistry was performed on two series of the PFC, staining every 1 in 7 free-floating sections to investigate calcium-binding proteins (calbindin and parvalbumin). Primary antibodies were mouse-anti-parvalbumin (1:500, Sigma-Aldrich, MO, United States), and mouse-anti-calbindin (1:500, Sigma-Aldrich, Missouri, United States) and secondary antibody donkey-anti-mouse (1:500, Merck KGaA, Darmstadt, Germany). Sections were incubated overnight at 4^{o}C in blocking solution (0.1 M phosphate buffer; 0.1% BSA; 0.3% Triton; 5% donkey serum) containing the primary antibody. Following overnight incubation, sections were incubated at room temperature for 90 min in blocking solution with the secondary antibody followed incubation in avidin-biotin complex (1:800 A and B in 0.1 M phosphate buffer; 90 min at room temperature). Staining was visualized using 3,3'-diaminobenzidine as a chromogen. Finally, sections were mounted on gelatine coated slides, dehydrated, cleared in xylene, and coverslipped with Entellan (Merck KGaA, Darmstadt, Germany).

Immunolabeled neurons were counted in both hemispheres within each mounted cortical section using imaging software (NIS-elements software). Four regions of interest (ROI) were investigated, including two dorsal regions and two ventral regions (see section "Results" for location). The ROI were selected based on distance away from the skull, and hence the overlying coil (dorsal ROI 1 and 2 vs. ventral ROI 3 and 4) and as representing more superficial (ROI 1 and 4) vs. deeper layers of the cortex (ROI 2 and 3), though we did not attempt to only capture particular layers. The areas were selected based on landmarks which enabled the same criteria to be applied in each section and ensured avoidance of double counting. The counting area was set to 0.24 mm^2 using a 20X objective lens, with images captured using a Nikon Eclipse 80i and photomicrographs taken with an integrated Digital Sight Camera (NIS-elements software). Immunostained neurons were only counted when soma and at least two dendrites were visible.

Statistical Analysis

Statistical analysis was performed using GraphPad Prism. Normality was checked with Shapiro–Wilks normality test. RMS startle response data was compared using a Mann–Whitney test. Group behavioral data (GPIAS), immunohistochemistry, and CAP thresholds were analyzed using two-way repeated measures analyses of variance (ANOVA) with associated Turkey's *post hoc* tests. Spontaneous and burst firing frequency between regions were analyzed using Kruskal–Wallis test with Dunn's multiple comparisons as the data were non-parametric.

RESULTS

The present study aimed to compare the effects of two common clinical rTMS protocols with sham treatment applied on PFC, in a guinea pig tinnitus model. The experimental design is shown in **Figure 2**. Baseline behavioral tinnitus testing was performed to establish that the neural circuitry underlying startle response and PPI was normal and to ensure there was no evidence of pre-existing tinnitus. Then surgery was performed to allow for measurement of auditory thresholds and to allow for exposure to a unilateral AT which caused permanent hearing loss. After a 1-week recovery period from surgery, animals resumed weekly behavioral testing for tinnitus. Once animals presented with behavioral signs of tinnitus, active rTMS (either 20 Hz or iTBS) or sham rTMS was administered over the PFC of the awake animal for 10 min, weekdays only, over a period of 2 weeks. Three days following cessation of this treatment, behavioral signs of tinnitus were reassessed, after which spontaneous firing of single neurons in the MGN was recorded under deep anesthesia. Brain tissue was subsequently obtained for immunohistochemical analysis of calcium-binding proteins in PFC.

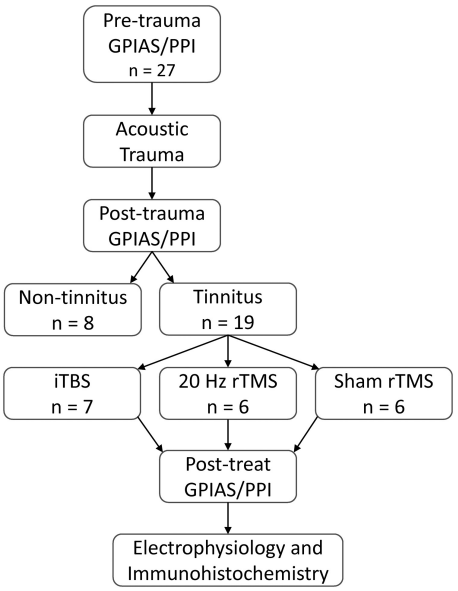

FIGURE 2 | Experimental overview. All animals tested for baseline GPIAS and PPI ensuring no pre-existing behavioral evidence of tinnitus and normal startle circuitry. Baseline behavioral tests were followed by an AT to induce hearing loss. Following hearing loss, animals were tested weekly for behavioural signs of tinnitus. Tinnitus animals were treated with active rTMS (20 Hz or iTBS) or sham rTMS administered over the PFC for 10 min daily, over 2 weeks (Monday to Friday). Following treatment, behavioral signs of tinnitus were assessed, along with spontaneous and burst firing rates of single neurons in the MGN, and subsequent post-euthanasia immunohistochemical analysis of calcium-binding proteins in the PFC.

Electromagnetic Field Modelling

Using 9.4 T MRI for 3D head imaging, these computer models were then used to investigate E-field spread and intensity (**Figure 1**). The 99th percentile highest E-field value produced in the cortex with the Cool-40 Rat coil set to 23% machine stimulus output was 39.5 V/m, 0.457 cm^3 of this brain volume contained an E-field of 36.23 V/m or above and 1.75 cm^3 measured an E-field of 24.15 V/m or above. The maximum E-field value found

in the thalamus was 20.4 V/m at a depth of 8.17 mm from the cortex surface and 99% of the E-field found in the thalamus was at 13.1 V/m or below. E-fields are shown in **Figure 1B**.

Behavioral Tests for Tinnitus

Before AT all animals showed significant suppression in the PPI test (**Figure 3A**; Pre-trauma data) as well as in the GPIAS test (**Figure 3B**; Pre-trauma data), both averaging approximately between 40 and 50%. When tinnitus develops, animals fail GPIAS (a lack of significant suppression between no gap and gap trials within an animal). Although the underlying mechanism of this phenomenon is still under debate (Eggermont, 2013, 2016; Galazyuk and Hébert, 2015; Salloum et al., 2016) the GPIAS test has been verified against psychophysical tests for tinnitus in animals (Turner et al., 2006) and is supported by some data from the human tinnitus population (Fournier and Hébert, 2013, 2020; Duda et al., 2020), although others have debated its use (Campolo et al., 2013; Morse and Vander Werff, 2019). As a lack of significant suppression in the GPIAS test could be due to hearing loss, i.e., the animals not being able to hear the background noise and hence not detecting the gap, or alternatively an issue with startle circuitry, PPI is implemented alongside the GPIAS to exclude these possibilities. Following AT all animals still showed significant suppression of PPI (**Figure 3A**) indicating that startle circuitry was working normally and that prepulse detection was not significantly affected by ipsilateral hearing loss. Mean PPI suppression was significantly different between the time-points ($F_{2,16} = 0.1733$, $p = 0.8424$) and the interaction between treatment and timepoint was non-significant ($F_{2,16} = 1.751$, $p = 0.2052$). The increase in PPI following AT may be an indication of hyperacusis being experienced by the animals (Chen et al., 2013; Thomas et al., 2019). Significant PPI suppression indicates that the background noise used in GPIAS should be detectable and a decreased suppression in GPIAS test is not due to their ipsilateral hearing loss. Therefore, a loss of significant suppression in just GPIAS would represent tinnitus.

Nineteen (15 males and 4 females) of the 27 animals developed a GPIAS deficit indicating behavioral signs of tinnitus. The eight non-tinnitus animals consisted of five males and three females and are not further discussed. Tinnitus development occurred between 3 and 16 weeks after AT (7.9 ± 0.8 weeks; mean ± SEM). Twelve of the 19 animals developed a GPIAS deficit at 8 kHz, four animals at 14 kHz, and three animals at both background frequencies. This is in line with our previous studies showing that the GPIAS deficit can occur at either or both frequencies (Mulders et al., 2014a, 2016, 2019; Leggett et al., 2018). The background noise/prepulse center frequencies used for GPIAS and PPI (8 and 14 kHz) were selected as these frequencies are just below the AT frequency in an audiogram region without threshold loss (8 kHz) and in a region just above the AT frequency showing threshold loss (14 kHz). Animals were randomly allocated to a treatment group with the 20 Hz consisting of five males and one female, iTBS groups consisting of six males and one female and the sham group of four males and two females. The average GPIAS suppression in the week of tinnitus development in each group is shown in **Figure 3B** (post-trauma data), demonstrating the dramatic and significant reduction in suppression (two-way repeated measures ANOVA, $F_{4,32} = 3.33$, $p = 0.0218$) as compared to before AT (pre-trauma data). *Post hoc* analysis showed that this significant change was present in all three groups: sham (Tukey's multiple comparisons test, $p < 0.001$); 20 Hz (Tukey's multiple comparisons test, $p < 0.001$); iTBS (Tukey's multiple comparisons test, $p < 0.05$).

After treatment with 20 Hz or iTBS, but not after sham rTMS, GPIAS suppression increased toward pre-AT levels and was significantly higher than at the post-trauma (time of tinnitus development) time-point (**Figure 3B**; Tukey multiple comparisons; sham, $p > 0.05$; 20 Hz, $p < 0.001$; iTBS, $p < 0.05$). This indicates that both the 20 Hz and iTBS treatment alleviated the signs of tinnitus. GPIAS suppression following sham treatment remained significantly lower than at pre-AT levels, and was unchanged from post-trauma levels, confirming the lack of effect of sham rTMS treatment.

CAP Audiograms

Compound action potential thresholds were measured immediately before and after AT as well as during the final electrophysiological experiment. This was done to ensure that (1) all animals had normal thresholds initially, (2) the AT caused similar effects in all groups, and (3) rTMS and sham treatments did not affect auditory thresholds. The pre-trauma CAP audiogram revealed no pre-existing threshold differences between groups (two-way ANOVA, $F_{20,160} = 0.94$, $p = 0.536$) and confirmed normal hearing in all groups (Johnstone et al., 1979). Immediately following AT animals showed a large temporary threshold loss from 6 to 24 kHz, with no significant differences between groups (**Figure 4A**; two-way ANOVA, $F_{20,160} = 0.45$, $p = 0.980$). At the time-point of the final electrophysiological recordings, CAP thresholds in the non-AT ear were normal in all groups (data not shown), but all animals showed a permanent threshold loss in the AT ear (**Figure 4B**), consistent with our previous guinea pig studies (Mulders and Robertson, 2009, 2011; Mulders et al., 2011). Threshold loss was not significantly different between groups (two-way ANOVA, $F_{20,160} = 0.37$, $p = 0.994$), which implies that any differences with regards to tinnitus outcomes, electrophysiology or immunohistochemistry are not due to differences in peripheral thresholds.

Firing Rates and Patterns in the MGN

Spontaneous firing data were collected from a total of 891 neurons (272 from sham rTMS, 295 from 20 Hz rTMS, and 324 from iTBS treated animals) within the right MGN, contralateral to the AT ear, as this represents the main auditory pathway. Spontaneous firing data are non-parametric and are shown as median with associated interquartile range (IQR), there was a significant difference between groups (**Figure 5A**; Kruskal–Wallis, $p < 0.0001$). MGN neurons from 20 Hz rTMS treated animals had a significantly lower firing rate (0.3 [1.2 IQR] spikes/s) compared to sham rTMS (0.7 [1.2 IQR]; Dunn's multiple comparisons test, $p < 0.0286$). Whereas, iTBS treated animals had a significantly higher firing rate (1.2 [1.9 IQR] spikes/s) compared to sham rTMS (Dunn's multiple comparisons test, $p < 0.0001$) and 20 Hz rTMS treated animals (Dunn's multiple comparisons test, $p < 0.0001$).

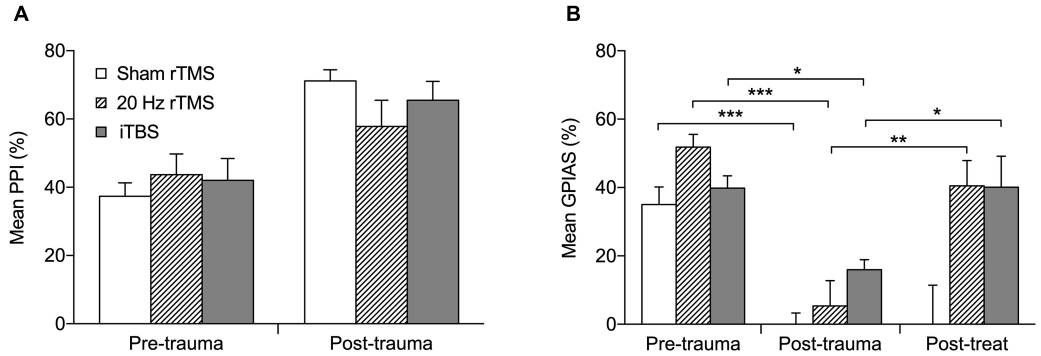

FIGURE 3 | Active rTMS ameliorates behavioral signs of tinnitus. **(A)** The percentage PPI at the frequency of tinnitus development. Mean percentage PPI is significantly different between pre-trauma and post-trauma time points ($p < 0.0001$). **(B)** The percentage GPIAS at the frequency of tinnitus development. For the three animals that developed GPIAS deficit at both 8 and 14 kHz, the average PPI and GPIAS for both frequencies was calculated. Graphs show the data before acoustic trauma (pre-trauma), at the time-point of tinnitus development (post-trauma), and after treatment on the day of final electrophysiological recordings (post-treat; GPIAS data only) for Sham rTMS ($n = 6$), 20 Hz rTMS ($n = 6$), iTBS ($n = 7$). All data mean + SEM, two-way repeated measures ANOVA with Tukey multiple comparisons; $^*p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

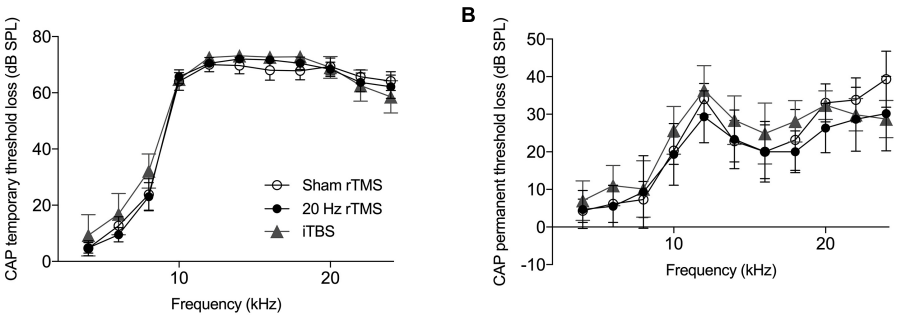

FIGURE 4 | Hearing threshold loss following acoustic trauma. Mean compound action potential threshold loss in the left ear. **(A)** Average CAP threshold loss measured at multiple frequencies (4–24 kHz) immediately following exposure to acoustic trauma. **(B)** Average CAP threshold loss at multiple frequencies (4–24 kHz) at the day of final electrophysiological recordings. All data mean ± SEM. No significant differences between groups identified.

Neurons were further analyzed on the basis of CF and pure tone response. Neurons that were pure tone responsive were subdivided into a CF < 10 kHz and ≥10 kHz. This was done as 10 kHz was used as the AT frequency and CAP threshold loss only occurred over and not below the AT frequency. The third group consisted of neurons that responded to noise, but not a pure tone (no CF data). Additionally, differentiated neurons that were just isolated offline based only on wave height and width were excluded from this further analysis. Differentiating neurons in this way indicated a significant difference in spontaneous firing rate between treatment groups (**Figure 5B**; Kruskal–Wallis, $p < 0.0001$). MGN neurons < 10 kHz from iTBS treated animals had a significantly higher firing rate (0.9[1.8 IQR] spikes/s) compared to sham rTMS (0.5 [1.1 IQR]; Dunn's multiple comparisons test, $p < 0.0001$) and 20 Hz treated animals (0.35 [1.2 IQR]; Dunn's multiple comparisons test, $p = 0.0007$). MGN neurons ≥ 10 kHz had a significantly different firing rate between iTBS (1.2 [2.3 IQR] spikes/sec) and 20 Hz rTMS (0.2 [0.75 IQR] Dunn's multiple comparisons test, $p < 0.0001$), however, active treatment groups were not significantly different from sham treated animals (0.8 [1.0 IQR]; Dunn's multiple comparisons test, $p > 0.05$). Similarly, the MGN neurons without CF were significantly different between iTBS treated animals (1.3 [2.1 IQR]) and 20 Hz (0.7 [1.5 IQR]; Dunn's multiple comparisons test, $p = 0.0018$), but not sham treatment (0.9 [4.3 IQR]; Dunn's multiple comparisons test, $p > 0.05$).

Burst analysis was performed only on neurons with a firing rate > 1 spike/s, resulting in burst analysis on 107 neurons from sham rTMS, 86 neurons from 20 Hz rTMS, and 174 from iTBS treated animals (**Figures 5C,E**). Burst rates (bursts per minute) were significantly different between the active rTMS treatments (Dunn's multiple comparisons test, $p = 0.0119$), however there were no significant differences between the active rTMS treatments and sham rTMS (Dunn's multiple comparisons test, $p > 0.05$) (**Figure 5C**; Kruskal–Wallis, $p = 0.0150$). Differentiating MGN neurons on the basis of CF (<10 kHz, ≥ 10 kHz or no CF) indicated no significant differences in burst firing rate between treatment groups (**Figure 5D**; Kruskal–Wallis, $p > 0.05$).

Mean burst duration was slightly but significantly longer after 20 Hz rTMS treatment (9.36 [2.19 IQR] ms), compared to sham rTMS (8.38 [1.29 IQR] ms; Dunn's multiple comparisons test, $p < 0.0001$) and iTBS treatments (8.65 [1.36 IQR] ms; Dunn's

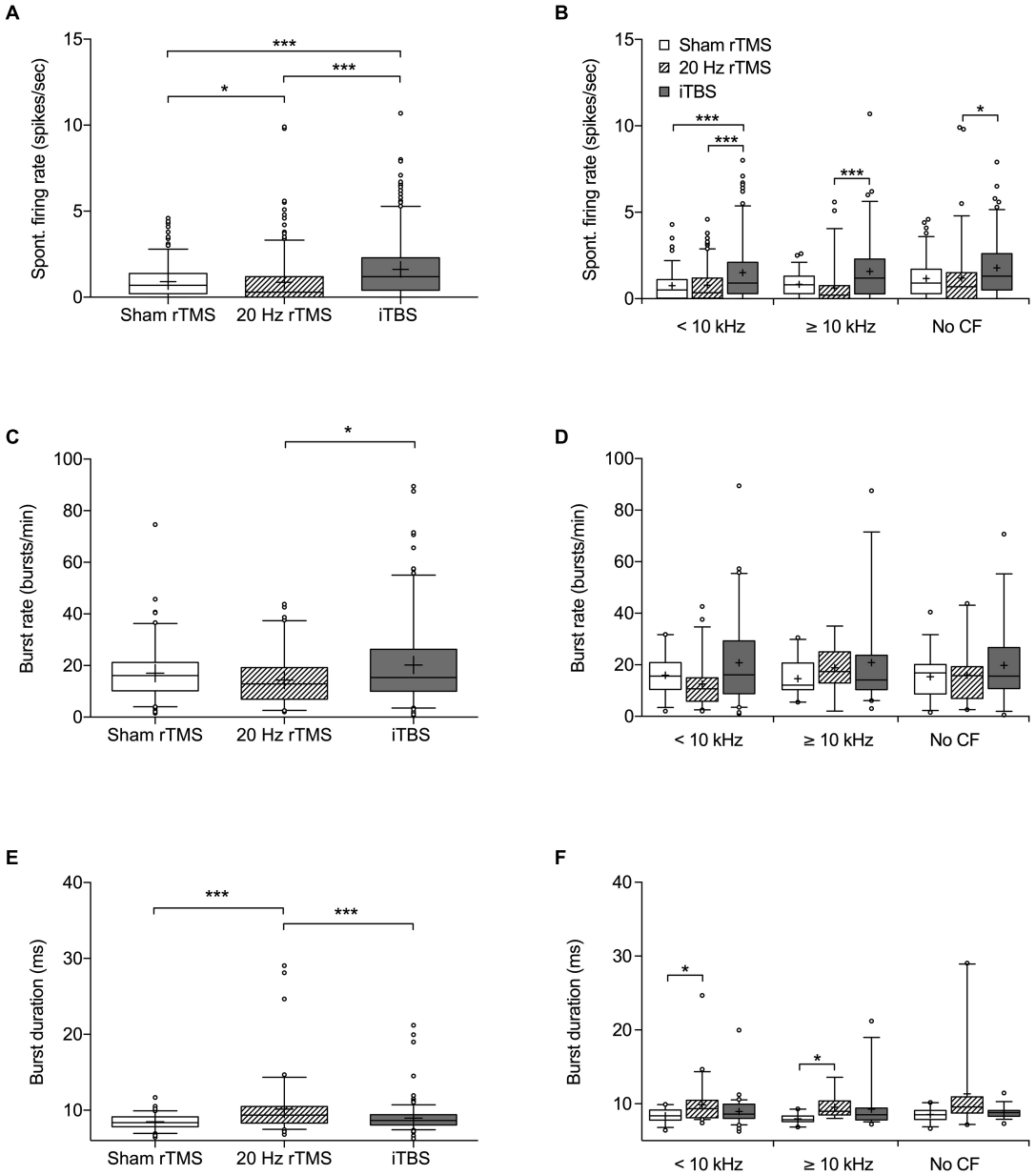

FIGURE 5 | Spontaneous firing rate and burst firing parameters of MGN neurons following rTMS. Box and whisker plot showing **(A)** spontaneous firing rates based on 272 neurons from sham, 295 neurons from 20 Hz rTMS, and 324 neurons from iTBS group. **(B)** Spontaneous firing rate of neurons differentiated into CF groups [< 10 kHz, ≥ 10 kHz, and No CF]. **(C)** Burst firing rate in bursts per minute based on 107 neurons from sham, 86 neurons from 20 Hz rTMS, and 174 neurons from iTBS. **(D)** Burst firing rate of neurons differentiated into CF groups [<10 kHz, ≥10 kHz, and No CF]. **(E)** Burst duration based on 107 neurons from sham, 86 neurons from 20 Hz rTMS, and 174 neurons from iTBS. **(F)** Burst duration of neurons differentiated into CF groups [<10 kHz, ≥10 kHz, and No CF]. All data median with box spanning interquartile range, minimum, maximum and outliers (>95%). Mean is indicated by +. Kruskal–Wallis with Dunn's multiple comparisons, $^{*}p < 0.05$, $^{***}p < 0.001$.

multiple comparisons test, $p = 0.0006$) (**Figure 5E**; Kruskal–Wallis, $p < 0.0001$). Differentiating neurons based on CF (<10 kHz, ≥10 kHz or no CF) indicated significant differences in burst duration between treatment groups (**Figure 5F**; Kruskal–Wallis, $p < 0.0001$). MGN neurons < 10 kHz from 20 Hz rTMS treated animals had a significantly longer burst duration (9.34 [2.35 IQR]) compared to sham rTMS (8.38 [1.137 IQR]; Dunn's multiple comparisons test, $p = 0.0315$), but not iTBS treated animals (8.61 [1.9 IQR]). Similarly, MGN neurons ≥ 10 kHz from 20 Hz rTMS treated animals were significantly longer (8.98 [1.88 IQR]) than sham rTMS (7.86 [0.78 IQR]), but not iTBS treated animals (8.52 [1.62 IQR]). MGN neurons that responded to noise, but not a pure tone (no CF data) were not significantly different between treatment groups.

Immunohistochemistry

TMS treatments have been shown to alter the densities of calcium-binding proteins in the cortex (Benali et al., 2011; Volz et al., 2013; Castillo−Padilla and Funke, 2016; Makowiecki et al., 2018; Mulders et al., 2019). In agreement, the densities of calbindin-positive neurons were significantly different between groups (two-way repeated measures ANOVA, $F_{6,48} = 2.5$, $p = 0.0360$). This significant increased density compared to sham treatment was present in ROI 1 (the most dorsal ROI, see **Figure 6A**) after both the 20 Hz rTMS (Sidak's multiple comparisons, $p = 0.0020$) and iTBS treatment (Sidak's multiple comparisons, $p = 0.0027$), but the other ROIs did not show a statistically significant difference (**Figure 6B**). The densities of parvalbumin-positive neurons were not significantly different (**Figure 6C**; two-way repeated measures ANOVA, $F_{6,48} = 1.1$, $p = 0.4008$).

DISCUSSION

The present study examined the effects of two excitatory rTMS protocols, commonly used in clinical settings, applied over PFC in an animal model of tinnitus. Both rTMS treatments resulted in a significant decrease in the behavioral signs of tinnitus compared to sham treatment. An increase in the density of calbindin within the most dorsal region of the PFC confirmed local effects of rTMS on the targeted area. Interestingly, the attenuation of tinnitus was accompanied by significant changes in spontaneous neuronal firing rate in opposite directions following the active rTMS protocols. Burst rate in the MGN was significantly different between 20 Hz rTMS and iTBS treatment, but there was only a significant change in burst duration compared to sham following 20 Hz rTMS. Collectively, the present findings provide support for the use of rTMS of the PFC to ameliorate tinnitus in humans, and lend support to current hypotheses implicating the PFC in the generation of tinnitus (Rauschecker et al., 2010, 2015). However, our data also suggest that the therapeutic benefit of stimulating PFC may not always be directly associated with inhibition of thalamic hyperactivity.

A key element of our experimental design was that the thalamus should not be directly stimulated by rTMS, but rather be indirectly activated by stimulation of the PFC. A study of the action potential thresholds of pyramidal cells in cortical slice preparations using stimulating electrodes suggested that ∼28 V/m was the minimum E-field intensity that could produce an action potential in some pyramidal neurons (Radman et al., 2009), although mean value was 57 ± 6 mV/mm for layer V/VI neurons and 81 ± 3 mV/mm for layer II/III pyramidal neurons. In addition, findings from computational models suggest that GABAergic interneurons have higher input resistance compared to pyramidal neurons (Hausser et al., 2000; Markram et al., 2004), and therefore may be even less susceptible to direct depolarization by the induced E-fields of TMS (Aberra et al., 2020). Modeling of the intensity of the E-field induced by our stimulation protocol confirmed that the thalamus received intensities well below the threshold for action potentials (∼20 V/m), while the superficial layers of the cortex

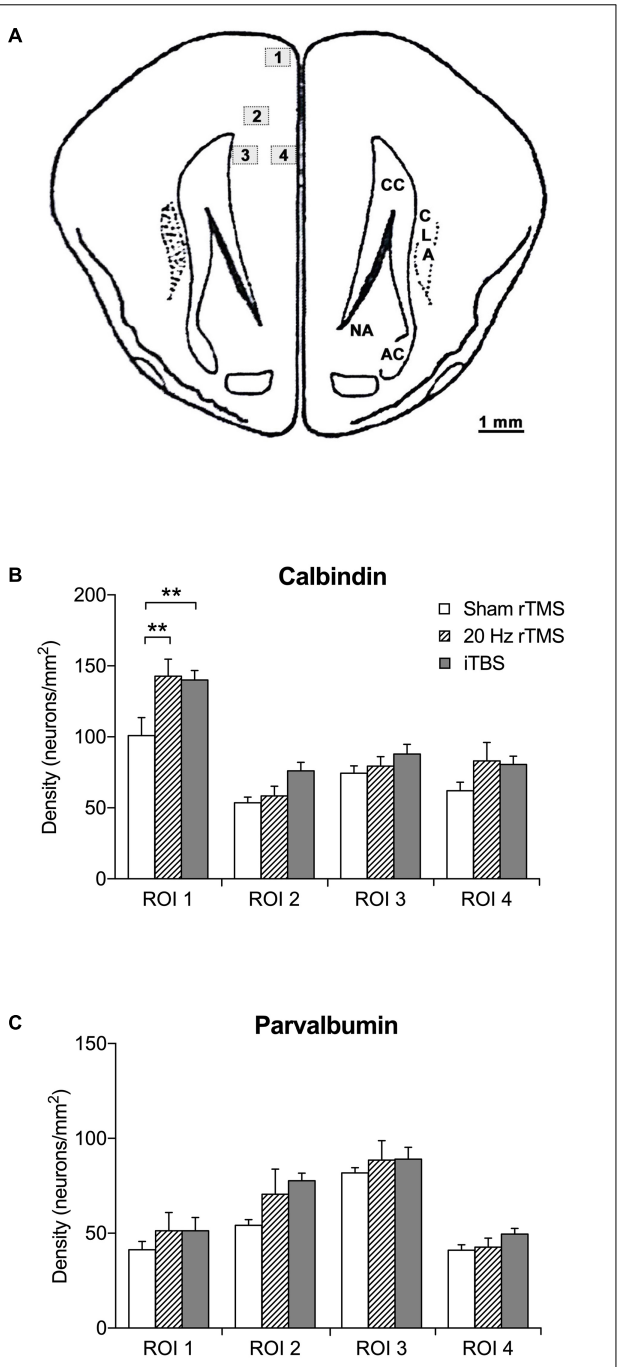

FIGURE 6 | Regions of interest and density of calbindin- and parvalbumin-positive neurons in the PFC. **(A)** Schematic illustration of location and area size of regions of interest (ROI). The counting frame area was set to 0.24 mm^2 using a X20 objective lens. The mean density per mm^2 was calculated by the mean number of calcium-binding immunoreactive neurons divided by the frame area (mm). Numbered boxed (1–4) represent locations and area (mm^2) of ROI. Image adapted from Rapisarda and Bacchelli (1977). **(B)** Mean density of calbindin-positive neurons in sham ($n = 6$), 20 Hz rTMS ($n = 6$), and iTBS ($n = 7$) treated animals within four ROI in the PFC. **(C)** Mean density of parvalbumin-positive neurons in sham ($n = 6$), 20 Hz rTMS ($n = 6$), and iTBS ($n = 7$) treated animals within four ROI in the PFC. All data mean + SEM, repeated measures ANOVA with Sidak's multiple comparisons; $^{**}p < 0.01$).

were stimulated at intensities within the lower range shown to elicit action potentials (up to 39.5 V/m). The E-field intensities at the 23% machine stimulus output used in this experiment are also below the motor threshold of 28% machine stimulus output using the same Cool-40 coil (Parthoens et al., 2016), determined using direct measurements of motor evoked potentials. Thus, our stimulation protocol likely induced action potentials in some pyramidal neurons, mostly in the superficial layers of the cortex, but did not directly activate thalamic neurons. However, E-field values below the threshold value required to produce action potentials have been observed inducing non-synaptic mechanisms of neuroplasticity such as structural reoganisation (Makowiecki et al., 2014), increased neurogenesis (Heath et al., 2018), gene regulation and increased intracellular calcium (Grehl et al., 2015; Clarke et al., 2017b), altered neuron excitability (Tang et al., 2016; Makowiecki et al., 2018), and glial changes (Clarke et al., 2017a; Cullen et al., 2019). Therefore, we cannot rule out that non-synaptic mechanisms of plasticity induced in the thalamus by low intensity magnetic fields may contribute to the changes in firing parameters that we report here.

The positive effects on tinnitus following both treatments, are in agreement with multiple clinical studies, which have reported beneficial effects on tinnitus following combined rTMS treatment over PFC and auditory cortex (Kleinjung et al., 2008; Lehner et al., 2013; Langguth et al., 2014; Ciminelli et al., 2020), most likely due to resulting inhibitory effects on auditory cortex. The mechanisms by which activation of PFC leads to attenuation of tinnitus remain to be elucidated. One possible mechanism would be that the observed amelioration of tinnitus is due to effects of PFC on the MGN and consequently on auditory cortex. Prior research has indicated direct and indirect projections from PFC to the thalamic reticular nucleus (Cornwall et al., 1990; Uylings and van Eden, 1991; Zikopoulos and Barbas, 2006) and the latter, in turn, has primarily inhibitory inputs to MGN (Yu et al., 2009). Hence, excitatory effects from our rTMS protocols on PFC (Pascual-Leone et al., 1994; Maeda et al., 2000; Huang et al., 2005; Pell et al., 2011) are thought to lead to activation in the thalamic reticular nucleus, which would result in inhibitory effects in MGN, which in turn would lead to inhibition of cortical activity.

However, despite the similar behavioral outcomes, with both active protocols ameliorating tinnitus, the two excitatory protocols affected activity in MGN differently: following 20 Hz rTMS, animals exhibited a decrease in spontaneous firing rate, no change in burst rate and a small yet significant increase in burst duration, whereas iTBS animals exhibited an increase in firing rate (both tonic and burst firing). The reason for these divergent effects after two excitatory protocols remains to be elucidated. The possibility exists that we recorded from different subdivisions in the MGN, however it is unlikely that there would be a consistent bias toward one subdivision in one group of animals and not in the other groups. It also seems unlikely that it is linked to the use of anesthesia, although our anesthesia protocol is likely to have resulted in lower overall spontaneous activity as we have discussed previously using our animal model (Cook et al., 2021).

The decreased activity that was observed in MGN after 20 Hz stimulation would be in line with the circuitry outlined above. However, the increased activity in the MGN following iTBS was unexpected: first, in view of the pathway outlined above and second, because an increase in firing rates in the auditory pathway is associated with the presence of tinnitus (Kalappa et al., 2014; Wu et al., 2016). Indeed, we have shown in our animal model that there is an increased tonic spontaneous firing in the MGN of animals with tinnitus compared to animals without tinnitus (Cook et al., 2021) and hence we expected reduced spontaneous firing rates in the MGN following both active protocols alongside the attenuation of tinnitus.

One possibility to explain how the different rTMS protocols caused similar behavioral outcomes despite different changes in MGN activity is that stimulation of PFC ameliorated the behavioral signs of tinnitus after iTBS treatment independently of MGN, possibly via a direct pathway linking PFC and auditory cortex (Pandya and Kuypers, 1969; Petrides and Pandya, 1988; Winkowski et al., 2018). Indeed, human studies have shown that tinnitus is associated with a dysfunctional pathway from PFC to auditory cortex (Song et al., 2015) and it may be this pathway that is responsible for the therapeutic effects observed. Further studies measuring activity in auditory cortex rather than MGN, after PFC rTMS stimulation are required to investigate this possibility.

Alternatively, the changes in burst firing induced by the rTMS protocols may have an impact on tinnitus perception and maintenance. Thalamic burst firing has been postulated to play a role in salience in alert states and may be involved in pathological conditions such as phantom perceptions (Llinás et al., 1999; Sherman, 2001; Llinás and Steriade, 2006). However, contrasting theories exist regarding the role of burst firing. The thalamocortical dysrhythmia theory proposes that in cases of phantom perception, such as tinnitus, disfacilitation or increased inhibition in the MGN, produces low-threshold calcium bursts, causing abnormal oscillations between the thalamus and cortex (Llinás et al., 1999; De Ridder et al., 2015). In the cortex, this abnormal oscillation pattern is detectable as gamma oscillations, which can produce a conscious sensory percept (Llinás et al., 1999; van der Loo et al., 2009; Caspary and Llano, 2017). This theory is supported by the results of Kalappa et al. (2014) who showed increased spontaneous burst firing in animals with tinnitus after an AT compared to unexposed controls without tinnitus. However, interestingly, data arising from the research area of chronic pain show that increased burst firing has been associated with inhibition of pain (Kim et al., 2003; Cheong et al., 2008). Indeed, a recent study showed that inducing thalamic bursts, using optogenetic activation of the thalamic reticular nucleus neurons, resulted in down-regulation of nociceptive behavior in awake mice (LeBlanc et al., 2017). Therefore, the significantly higher burst firing (increased burst rate in iTBS and increased burst duration in 20 Hz rTMS) observed following rTMS treatment in the present study could potentially serve to prevent the transmission of the tinnitus signal to the cortex.

In addition, high frequency rTMS of the PFC has been shown to affect the neurochemical make-up of the mesolimbic system, modulating dopamine in the anterior cingulate gyrus in healthy humans (Cho and Strafella, 2009) and in the dorsal hippocampus, the shell of the nucleus accumbens and the dorsal striatum in rats (Keck et al., 2000). Structural and functional malfunction of

the mesolimbic system has been implicated in depression and chronic pain (Belujon and Grace, 2017; Serafini et al., 2020) and in addition, the circuitry thought to be involved in chronic pain and tinnitus shows remarkable overlap (Rauschecker et al., 2015). Hence, alteration of the mesolimbic system by rTMS stimulation may also be involved in the positive effects observed on tinnitus perception.

Even though the two rTMS protocols influenced thalamic activity in different ways, the effects on the density of the calcium-binding protein calbindin in the PFC were similar which is in line with both protocols being excitatory. Calbindin densities were increased in the most dorsal layers of the PFC only, which suggest a proximity-dependent relationship between density and the rTMS coil. The increase in calbindin neuron densities may be due to a reduction of protein degradation or increased protein synthesis. The increased density is in agreement with our previous study applying low-intensity rTMS stimulation to PFC (Mulders et al., 2019) and another study using rTMS over mouse visual cortex (Makowiecki et al., 2018). However, a study using 2 weeks of iTBS in rats showed reduced parvalbumin expression (Castillo−Padilla and Funke, 2016). This discrepancy may be due to species differences (Lensjø et al., 2017) or differences in baseline cortical activity (Makowiecki et al., 2018). How the increased density is related to PFC activity remains to be investigated. Parvalbumin and calbindin are both present in inhibitory neurons in the cortex (Druga, 2009; Atallah et al., 2012). In view of the excitatory protocols used (Pascual-Leone et al., 1994; Maeda et al., 2000; Huang et al., 2005; Pell et al., 2011), the higher density of immunopositive neurons could be the result of a homeostatic response to increased glutamatergic neuronal activity.

Finally, our data also suggest that the animals may have developed hyperacusis after the AT as the PPI showed a significant increase following AT, which has been suggested as a marker for hyperacusis (Chen et al., 2013; Thomas et al., 2019). Hyperacusis is a common co-morbidity in tinnitus patients (Schecklmann et al., 2014), as both are strongly associated with hearing loss (Baguley et al., 2013; Paulin et al., 2016). Recent studies in human subjects with tinnitus with or without hyperacusis suggest that hyperacusis is associated with increased neural gain which shown as an increase in sound evoked activity whereas tinnitus is associated with increased noise and reduced activity at the tinnitus frequency (Hofmeier et al., 2021; Koops and van Dijk, 2021). Unfortunately, in the present study, PPI was not measured after the rTMS treatments, which would be a useful additional measure in future studies to provide further information regarding the neural substrates underlying tinnitus and hyperacusis.

In summary, our study indicates that excitatory rTMS over PFC results in significant improvements in tinnitus. The mechanism behind this improvement needs further study to elucidate the exact circuitry.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusion of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The animal study was reviewed and approved by Animal Ethics Committee of The University of Western Australia.

AUTHOR CONTRIBUTIONS

WM and JR contributed to conception and design of the study. JZ and HT collected the data and performed the statistical analysis. KB contributed to data processing of electrophysiological single neuron recordings. KL contributed to data collection. JZ wrote the first draft of the manuscript. WM, JR, and SB wrote sections of the manuscript. SB produced guinea pig brain models and simulated induced electric fields. All the authors contributed to manuscript revision, read, and approved the submitted version.

FUNDING

This work was supported by funds provided by the School of Human Sciences, University of Western Australia, Crawley, WA, Australia. JR was supported by a fellowship from Multiple Sclerosis association of WA, and the Perron Institute for Neurological and Translational Science.

REFERENCES

Aberra, A. S., Wang, B., Grill, W. M., and Peterchev, A. V. (2020). Simulation of transcranial magnetic stimulation in head model with morphologically-realistic cortical neurons. *Brain Stimulat.* 13, 175–189. doi: 10.1016/j.brs.2019.10.002

Anderson, L., Malmierca, M., Wallace, M., and Palmer, A. (2006). Evidence for a direct, short latency projection from the dorsal cochlear nucleus to the auditory thalamus in the guinea pig. *European Journal of Neuroscience* 24, 491–498. doi: 10.1111/j.1460-9568.2006.04930.x

Atallah, B. V., Bruns, W., Carandini, M., and Scanziani, M. (2012). Parvalbumin-expressing interneurons linearly transform cortical responses to visual stimuli. *Neuron* 73, 159–170. doi: 10.1016/j.neuron.2011.12.013

Attene, M. (2010). A lightweight approach to repairing digitized polygon meshes. *The visual computer* 26, 1393–1406. doi: 10.1007/s00371-010-0416-3

Baguley, D., McFerran, D., and Hall, D. (2013). Tinnitus. *Lancet* 382, 1600–1607. doi: 10.1016/S0140-6736(13)60142-7

Barry, K., Paolini, A., Robertson, D., and Mulders, W. (2015). Modulation of medial geniculate nucleus neuronal activity by electrical stimulation of the nucleus accumbens. *Neuroscience* 308, 1–10. doi: 10.1016/j.neuroscience.2015.09.008

Barry, K., Robertson, D., and Mulders, W. (2019). Changes in auditory thalamus neural firing patterns after acoustic trauma in rats. *Hearing research* 379, 89–97. doi: 10.1016/j.heares.2019.05.001

Barry, K. M., Robertson, D., and Mulders, W. (2017). Medial geniculate neurons show diverse effects in response to electrical stimulation of prefrontal cortex. *Hear Res* 353, 204–212. doi: 10.1016/j.heares.2017.07.002

Bartlett, E. L. (2013). The organization and physiology of the auditory thalamus and its role in processing acoustic features important for speech perception. *Brain and language* 126, 29–48. doi: 10.1016/j.bandl.2013.03.003

Basura, G. J., Koehler, S. D., and Shore, S. E. (2015). Bimodal stimulus timing-dependent plasticity in primary auditory cortex is altered after noise exposure with and without tinnitus. *Journal of neurophysiology* 114, 3064–3075. doi: 10.1152/jn.00319.2015

Belujon, P., and Grace, A. A. (2017). Dopamine system dysregulation in major depressive disorders. *International Journal of Neuropsychopharmacology* 20, 1036–1046. doi: 10.1093/ijnp/pyx056

Benali, A., Trippe, J., Weiler, E., Mix, A., Petrasch-Parwez, E., Girzalsky, W., et al. (2011). Theta-burst transcranial magnetic stimulation alters cortical inhibition. *Journal of Neuroscience* 31, 1193–1203. doi: 10.1523/jneurosci.1379-10.2011

Bicks, L. K., Koike, H., Akbarian, S., and Morishita, H. (2015). Prefrontal cortex and social cognition in mouse and man. *Frontiers in psychology* 6:1805.

Brotherton, H., Plack, C. J., Maslin, M., Schaette, R., and Munro, K. J. (2015). Pump up the volume: could excessive neural gain explain tinnitus and hyperacusis? *Audiology and Neurotology* 20, 273–282. doi: 10.1159/000430459

Campolo, J., Lobarinas, E., and Salvi, R. (2013). Does tinnitus "fill in" the silent gaps? *Noise & health* 15,

Caspary, D. M., and Llano, D. A. (2017). Auditory thalamic circuits and GABAA receptor function: Putative mechanisms in tinnitus pathology. *Hear Res* 349, 197–207. doi: 10.1016/j.heares.2016.08.009

Castillo−Padilla, D. V., and Funke, K. (2016). Effects of chronic iTBS−rTMS and enriched environment on visual cortex early critical period and visual pattern discrimination in dark−reared rats. *Developmental neurobiology* 76, 19–33. doi: 10.1002/dneu.22296

Chen, G., Lee, C., Sandridge, S., Butler, H., Manzoor, N., and Kaltenbach, J. (2013). Behavioral evidence for possible simultaneous induction of hyperacusis and tinnitus following intense sound exposure. *Journal of the Association for Research in Otolaryngology* 14, 413–424. doi: 10.1007/s10162-013-0375-2

Cheong, E., Lee, S., Choi, B. J., Sun, M., Lee, C. J., and Shin, H. S. (2008). Tuning thalamic firing modes via simultaneous modulation of T- and L-type Ca2+ channels controls pain sensory gating in the thalamus. *J Neurosci* 28, 13331–13340. doi: 10.1523/JNEUROSCI.3013-08.2008

Cho, S. S., and Strafella, A. P. (2009). rTMS of the left dorsolateral prefrontal cortex modulates dopamine release in the ipsilateral anterior cingulate cortex and orbitofrontal cortex. *PloS one* 4:e6725. doi: 10.1371/journal.pone.0006725

Ciminelli, P., Machado, S., Palmeira, M., Coutinho, E. S. F., Sender, D., and Nardi, A. E. (2020). Dorsomedial Prefrontal Cortex Repetitive Transcranial Magnetic Stimulation for Tinnitus: Promising Results of a Blinded, Randomized, Sham-Controlled Study. *Ear and Hearing*

Clarke, D., Penrose, M. A., Harvey, A. R., Rodger, J., and Bates, K. A. (2017a). Low intensity rTMS has sex-dependent effects on the local response of glia following a penetrating cortical stab injury. *Experimental neurology* 295, 233–242. doi: 10.1016/j.expneurol.2017.06.019

Clarke, D., Penrose, M. A., Penstone, T., Fuller-Carter, P. I., Hool, L. C., Harvey, A. R., et al. (2017b). Frequency-specific effects of repetitive magnetic stimulation on primary astrocyte cultures. *Restorative neurology and neuroscience* 35, 557–569. doi: 10.3233/rnn-160708

Cook, J., Barry, K., Zimdahl, J., Leggett, K., and Mulders, W. (2021). Spontaneous firing patterns in the medial geniculate nucleus in a guinea pig model of tinnitus. *Hearing Research* 403, 108190. doi: 10.1016/j.heares.2021.108190

Cornwall, J., Cooper, J., and Phillipson, O. (1990). Projections to the rostral reticular thalamic nucleus in the rat. *Experimental brain research* 80, 157–171.

Crippa, A., Lanting, C. P., Van Dijk, P., and Roerdink, J. B. (2010). A diffusion tensor imaging study on the auditory system and tinnitus. *The open neuroimaging journal* 4, 16. doi: 10.2174/1874440001004010016

Cullen, C. L., Senesi, M., Tang, A. D., Clutterbuck, M. T., Auderset, L., O'Rourke, M. E., et al. (2019). Low−intensity transcranial magnetic stimulation promotes the survival and maturation of newborn oligodendrocytes in the adult mouse brain. *Glia* 67, 1462–1477. doi: 10.1002/glia.23620

De Ridder, D., Song, J.-J., and Vanneste, S. (2013). Frontal cortex TMS for tinnitus. *Brain stimulation* 6, 355–362. doi: 10.1016/j.brs.2012.07.002

De Ridder, D., Vanneste, S., Langguth, B., and Llinas, R. (2015). Thalamocortical dysrhythmia: a theoretical update in tinnitus. *Frontiers in neurology* 6:124.

De Ridder, D., Vanneste, S., Weisz, N., Londero, A., Schlee, W., Elgoyhen, A. B., et al. (2014). An integrative model of auditory phantom perception: tinnitus as a unified percept of interacting separable subnetworks. *Neurosci Biobehav Rev* 44, 16–32. doi: 10.1016/j.neubiorev.2013.03.021

Druga, R. (2009). Neocortical inhibitory system. *Folia Biol (Praha)* 55, 201–217.

Duda, V., Scully, O., Baillargeon, M.-S., and Hébert, S. (2020). Does tinnitus fill in the gap using electrophysiology? A scoping review. *Otolaryngologic Clinics of North America* 53, 563–582. doi: 10.1016/j.otc.2020.03.006

Eggermont, J. J. (2013). Hearing loss, hyperacusis, or tinnitus: what is modeled in animal research? *Hearing research* 295, 140–149. doi: 10.1016/j.heares.2012.01.005

Eggermont, J. J. (2016). Can animal models contribute to understanding tinnitus heterogeneity in humans? *Frontiers in aging neuroscience* 8:265.

Eggermont, J. J., and Roberts, L. E. (2015). Tinnitus: animal models and findings in humans. *Cell Tissue Res* 361, 311–336. doi: 10.1007/s00441-014-1992-8

Fournier, P., and Hébert, S. (2013). Gap detection deficits in humans with tinnitus as assessed with the acoustic startle paradigm: does tinnitus fill in the gap? *Hearing Research* 295, 16–23. doi: 10.1016/j.heares.2012.05.011

Fournier, P., and Hébert, S. (2020). The gap prepulse inhibition of the acoustic startle (GPIAS) paradigm to assess auditory temporal processing: Monaural versus binaural presentation. *Psychophysiology* e13755.

Galazyuk, A., and Hébert, S. (2015). Gap-prepulse inhibition of the acoustic startle reflex (GPIAS) for tinnitus assessment: current status and future directions. *Frontiers in neurology* 6:88.

Geuzaine, C. (2007). "GetDP: a general finite−element solver for the de Rham complex," in *PAMM: Proceedings in Applied Mathematics and Mechanics*, (Wiley Online Library), 1010603–1010604. doi: 10.1002/pamm.200700750

Grehl, S., Viola, H. M., Fuller-Carter, P. I., Carter, K. W., Dunlop, S. A., Hool, L. C., et al. (2015). Cellular and molecular changes to cortical neurons following low intensity repetitive magnetic stimulation at different frequencies. *Brain stimulation* 8, 114–123. doi: 10.1016/j.brs.2014.09.012

Guinan, J. J. Jr. (2018). Olivocochlear efferents: Their action, effects, measurement and uses, and the impact of the new conception of cochlear mechanical responses. *Hearing research* 362, 38–47. doi: 10.1016/j.heares.2017.12.012

Guo, W., Clause, A. R., Barth-Maron, A., and Polley, D. B. (2017). A corticothalamic circuit for dynamic switching between feature detection and discrimination. *Neuron* 95, 180.–194.

Hausser, M., Spruston, N., and Stuart, G. J. (2000). Diversity and dynamics of dendritic signaling. *Science* 290, 739–744. doi: 10.1126/science.290.5492.739

Heath, A., Lindberg, D. R., Makowiecki, K., Gray, A., Asp, A. J., Rodger, J., et al. (2018). Medium-and high-intensity rTMS reduces psychomotor agitation with distinct neurobiologic mechanisms. *Translational psychiatry* 8, 1–13.

Hofmeier, B., Wertz, J., Refat, F., Hinrichs, P., Saemisch, J., Singer, W., et al. (2021). Functional biomarkers that distinguish between tinnitus with and without hyperacusis. *Clinical and translational medicine* 11,

Homma, N. Y., Happel, M. F., Nodal, F. R., Ohl, F. W., King, A. J., and Bajo, V. M. (2017). A role for auditory corticothalamic feedback in the perception of complex sounds. *Journal of Neuroscience* 37, 6149–6161. doi: 10.1523/jneurosci.0397-17.2017

Hoppenrath, K., and Funke, K. (2013). Time-course of changes in neuronal activity markers following iTBS-TMS of the rat neocortex. *Neuroscience letters* 536, 19–23. doi: 10.1016/j.neulet.2013.01.003

Huang, Y.-Z., Edwards, M. J., Rounis, E., Bhatia, K. P., and Rothwell, J. C. (2005). Theta burst stimulation of the human motor cortex. *Neuron* 45, 201–206.

Jenkinson, M., Beckmann, C. F., Behrens, T. E., Woolrich, M. W., and Smith, S. M. (2012). Fsl. *Neuroimage* 62, 782–790.

Johnstone, J., Alder, V., Johnstone, B., Robertson, D., and Yates, G. (1979). Cochlear action potential threshold and single unit thresholds. *The Journal of the Acoustical Society of America* 65, 254–257. doi: 10.1121/1.382244

Kalappa, B. I., Brozoski, T. J., Turner, J. G., and Caspary, D. M. (2014). Single unit hyperactivity and bursting in the auditory thalamus of awake rats directly correlates with behavioural evidence of tinnitus. *The Journal of physiology* 592, 5065–5078. doi: 10.1113/jphysiol.2014.278572

Keck, M. E., Sillaber, I., Ebner, K., Welt, T., Toschi, N., Kaehler, S. T., et al. (2000). Acute transcranial magnetic stimulation of frontal brain regions selectively modulates the release of vasopressin, biogenic amines and amino acids in the rat brain. *European Journal of Neuroscience* 12, 3713–3720. doi: 10.1046/j.1460-9568.2000.00243.x

Keifer, O. P. Jr., Gutman, D. A., Hecht, E. E., Keilholz, S. D., and Ressler, K. J. (2015). A comparative analysis of mouse and human medial geniculate nucleus connectivity: a DTI and anterograde tracing study. *NeuroImage* 105, 53–66. doi: 10.1016/j.neuroimage.2014.10.047

Kepecs, A., and Lisman, J. (2004). How to read a burst duration code. *Neuroscomputing* 58-60, 1–6. doi: 10.1016/j.neucom.2004.01.014

Kim, D., Park, D., Choi, S., Lee, S., Sun, M., Kim, C., et al. (2003). Thalamic control of visceral nociception mediated by T-type Ca2+ channels. *Science* 302, 117–119. doi: 10.1126/science.1088886

Kimura, A., and Imbe, H. (2015). Anatomically structured burst spiking of thalamic reticular nucleus cells: implications for distinct modulations of sensory processing in lemniscal and non-lemniscal thalamocortical loop circuitries. *Eur J Neurosci* 41, 1276–1293. doi: 10.1111/ejn.12874

Kleinjung, T., Eichhammer, P., Landgrebe, M., Sand, P., Hajak, G., Steffens, T., et al. (2008). Combined temporal and prefrontal transcranial magnetic stimulation for tinnitus treatment: a pilot study. *Otolaryngology—Head and Neck Surgery* 138, 497–501. doi: 10.1016/j.otohns.2007.12.022

Koops, E., and van Dijk, P. (2021). Hyperacusis in tinnitus patients relates to enlarged subcortical and cortical responses to sound except at the tinnitus frequency. *Hear Res* 401, 108158. doi: 10.1016/j.heares.2020.108158

Labedi, A., Benali, A., Mix, A., Neubacher, U., and Funke, K. (2014). Modulation of inhibitory activity markers by intermittent theta-burst stimulation in rat cortex is NMDA-receptor dependent. *Brain stimulation* 7, 394–400. doi: 10.1016/j.brs.2014.02.010

Langguth, B., Landgrebe, M., Frank, E., Schecklmann, M., Sand, P., Vielsmeier, V., et al. (2014). Efficacy of different protocols of transcranial magnetic stimulation for the treatment of tinnitus: pooled analysis of two randomized controlled studies. *The World Journal of Biological Psychiatry* 15, 276–285. doi: 10.3109/15622975.2012.708438

Laubach, M., Amarante, L. M., Swanson, K., and White, S. R. (2018). What, if anything, is rodent prefrontal cortex? *Eneuro* 5,

LeBlanc, B. W., Cross, B., Smith, K. A., Roach, C., Xia, J., Chao, Y.-C., et al. (2017). Thalamic bursts down-regulate cortical theta and nociceptive behavior. *Scientific reports* 7, 2482.

Leggett, K., Mendis, V., and Mulders, W. (2018). Divergent responses in the gap prepulse inhibition of the acoustic startle reflex in two different guinea pig colonies. *The international tinnitus journal* 22, 1–9.

Lehner, A., Schecklmann, M., Poeppl, T. B., Kreuzer, P. M., Vielsmeier, V., Rupprecht, R., et al. (2013). Multisite rTMS for the treatment of chronic tinnitus: stimulation of the cortical tinnitus network—a pilot study. *Brain topography* 26, 501–510. doi: 10.1007/s10548-012-0268-4

Lensjø, K. K., Christensen, A. C., Tennøe, S., Fyhn, M., and Hafting, T. (2017). Differential expression and cell-type specificity of perineuronal nets in hippocampus, medial entorhinal cortex, and visual cortex examined in the rat and mouse. *eneuro* 4,

Llinás, R. R., Ribary, U., Jeanmonod, D., Kronberg, E., and Mitra, P. P. (1999). Thalamocortical dysrhythmia: a neurological and neuropsychiatric syndrome characterized by magnetoencephalography. *Proceedings of the National Academy of Sciences* 96, 15222–15227. doi: 10.1073/pnas.96.26.15222

Llinás, R. R., and Steriade, M. (2006). Bursting of thalamic neurons and states of vigilance. *Journal of neurophysiology* 95, 3297–3308. doi: 10.1152/jn.00166.2006

Markram, H., Toledo-Rodriguez, M., Wang, Y., Gupta, A., Silberberg, G., and Wu, C. (2004). Interneurons of the neocortical inhibitory system. *Nat. Rev. Neurosci.* 5, 793–807. doi: 10.1038/nrn1519

Maeda, F., Keenan, J. P., Tormos, J. M., Topka, H., and Pascual-Leone, A. (2000). Modulation of corticospinal excitability by repetitive transcranial magnetic stimulation. *Clinical Neurophysiology* 111, 800–805. doi: 10.1016/s1388-2457(99)00323-5

Makowiecki, K., Garrett, A., Harvey, A. R., and Rodger, J. (2018). Low-intensity repetitive transcranial magnetic stimulation requires concurrent visual system activity to modulate visual evoked potentials in adult mice. *Scientific reports* 8, 1–13. doi: 10.1016/s0168-5597(97)00096-8

Makowiecki, K., Harvey, A. R., Sherrard, R. M., and Rodger, J. (2014). Low-intensity repetitive transcranial magnetic stimulation improves abnormal visual cortical circuit topography and upregulates BDNF in mice. *Journal of Neuroscience* 34, 10780–10792. doi: 10.1523/jneurosci.0723-14.2014

McCormick, D. A. (1992). Neurotransmitter actions in the thalamus and cerebral cortex and their role in neuromodulation of thalamocortical activity. *Progress in neurobiology* 39, 337–388. doi: 10.1016/0301-0082(92)90012-4

Mellott, J. G., Foster, N. L., Ohl, A. P., and Schofield, B. R. (2014). Excitatory and inhibitory projections in parallel pathways from the inferior colliculus to the auditory thalamus. *Frontiers in neuroanatomy* 8:124.

Mix, A., Benali, A., and Funke, K. (2014). Strain differences in the effect of rTMS on cortical expression of calcium-binding proteins in rats. *Experimental brain research* 232, 435–442. doi: 10.1007/s00221-013-3751-6

Morse, K., and Vander Werff, K. R. (2019). Comparison of silent gap in noise cortical auditory evoked potentials in matched tinnitus and no-tinnitus control subjects. *American journal of audiology* 28, 260–273. doi: 10.1044/2018_aja-18-0074

Mulders, W., Ding, D., Salvi, R., and Robertson, D. (2011). Relationship between auditory thresholds, central spontaneous activity, and hair cell loss after acoustic trauma. *Journal of Comparative Neurology* 519, 2637–2647. doi: 10.1002/cne.22644

Mulders, W., and Robertson, D. (2009). Hyperactivity in the auditory midbrain after acoustic trauma: dependence on cochlear activity. *Neuroscience* 164, 733–746. doi: 10.1016/j.neuroscience.2009.08.036

Mulders, W., and Robertson, D. (2011). Progressive centralization of midbrain hyperactivity after acoustic trauma. *Neuroscience* 192, 753–760. doi: 10.1016/j.neuroscience.2011.06.046

Mulders, W. H., Barry, K. M., and Robertson, D. (2014a). Effects of furosemide on cochlear neural activity, central hyperactivity and behavioural tinnitus after cochlear trauma in guinea pig. *PloS one* 9:e97948. doi: 10.1371/journal.pone.0097948

Mulders, W. H., Leggett, K., Mendis, V., Tarawneh, H., Wong, J., and Rodger, J. (2019). Low-intensity repetitive transcranial magnetic stimulation over prefrontal cortex in an animal model alters activity in the auditory thalamus but does not affect behavioural measures of tinnitus. *Experimental brain research* 237, 883–896. doi: 10.1007/s00221-018-05468-w

Mulders, W. H., Rodger, J., Yates, C. G., and Robertson, D. (2014b). Modulation of gene expression in guinea pig paraflocculus after induction of hearing loss. *F1000Research* 3.

Mulders, W. H. A. M., Vooys, V., Makowiecki, K., Tang, A., and Rodger, J. (2016). The effects of repetitive transcranial magnetic stimulation in an animal model of tinnitus. *Sci. Rep.* 6, 38234.

Norena, A., and Eggermont, J. (2003). Changes in spontaneous neural activity immediately after an acoustic trauma: implications for neural correlates of tinnitus. *Hearing research* 183, 137–153. doi: 10.1016/s0378-5955(03)00225-9

O'Donnell, P., Lavin, A., Enquist, L. W., Grace, A. A., and Card, J. P. (1997). Interconnected parallel circuits between rat nucleus accumbens and thalamus revealed by retrograde transynaptic transport of pseudorabies virus. *J Neurosci* 17, 2143–2167. doi: 10.1523/jneurosci.17-06-02143.1997

Pandya, D. N., and Kuypers, H. G. (1969). Cortico-cortical connections in the rhesus monkey. *Brain research* 13, 13–36. doi: 10.1016/0006-8993(69)90141-3

Parthoens, J., Verhaeghe, J., Servaes, S., Miranda, A., Stroobants, S., and Staelens, S. (2016). Performance characterization of an actively cooled repetitive transcranial magnetic stimulation coil for the rat. *Neuromodulation: Technology at the Neural Interface* 19, 459–468. doi: 10.1111/ner.12387

Pascual-Leone, A., Valls-Solé, J., Wassermann, E. M., and Hallett, M. (1994). Responses to rapid-rate transcranial magnetic stimulation of the human motor cortex. *Brain* 117, 847–858. doi: 10.1093/brain/117.4.847

Paulin, J., Andersson, L., and Nordin, S. (2016). Characteristics of hyperacusis in the general population. *Noise & health* 18, 178. doi: 10.4103/1463-1741.189244

Pell, G. S., Roth, Y., and Zangen, A. (2011). Modulation of cortical excitability induced by repetitive transcranial magnetic stimulation: influence of timing and geometrical parameters and underlying mechanisms. *Progress in neurobiology* 93, 59–98. doi: 10.1016/j.pneurobio.2010.10.003

Petrides, M., and Pandya, D. N. (1988). Association fiber pathways to the frontal cortex from the superior temporal region in the rhesus monkey. *Journal of Comparative Neurology* 273, 52–66. doi: 10.1002/cne.902730106

Radman, T., Ramos, R. L., Brumberg, J. C., and Bikson, M. (2009). Role of cortical cell type and morphology in subthreshold and suprathreshold uniform electric field stimulation in vitro. *Brain stimulation* 2, 215.–228.

Rapisarda, C., and Bacchelli, B. (1977). The brain of the guinea pig in stereotaxic coordinates. *Archivio di scienze biologiche* 61, 1–37. doi: 10.1007/978-1-4419-8372-5_1

Rauschecker, J. P., Leaver, A. M., and Mühlau, M. (2010). Tuning out the noise: limbic-auditory interactions in tinnitus. *Neuron* 66, 819–826. doi: 10.1016/j.neuron.2010.04.032

Rauschecker, J. P., May, E. S., Maudoux, A., and Ploner, M. (2015). Frontostriatal Gating of Tinnitus and Chronic Pain. *Trends Cogn Sci* 19, 567–578. doi: 10.1016/j.tics.2015.08.002

Robertson, D., Bester, C., Vogler, D., and Mulders, W. H. (2013). Spontaneous hyperactivity in the auditory midbrain: relationship to afferent input. *Hearing research* 295, 124–129. doi: 10.1016/j.heares.2012.02.002

Rolls, E. T. (2015). Limbic systems for emotion and for memory, but no single limbic system. *Cortex* 62, 119–157. doi: 10.1016/j.cortex.2013.12.005

Salloum, R., Sandridge, S., Patton, D., Stillitano, G., Dawson, G., Niforatos, J., et al. (2016). Untangling the effects of tinnitus and hypersensitivity to sound (hyperacusis) in the gap detection test. *Hearing research* 331, 92–100. doi: 10.1016/j.heares.2015.10.005

Saturnino, G. B., Puonti, O., Nielsen, J. D., Antonenko, D., Madsen, K. H., and Thielscher, A. (2019). "SimNIBS 2.1: a comprehensive pipeline for individualized electric field modelling for transcranial brain stimulation," in *Brain and Human Body Modeling*, (Springer), 3–25. doi: 10.1007/978-3-030-21293-3_1

Scannell, J., Burns, G., Hilgetag, C., O'Neil, M., and Young, M. P. (1999). The connectional organization of the cortico-thalamic system of the cat. *Cerebral Cortex* 9, 277–299. doi: 10.1093/cercor/9.3.277

Schecklmann, M., Landgrebe, M., Langguth, B., and Group, T. D. S. (2014). Phenotypic characteristics of hyperacusis in tinnitus. *PloS one* 9:e86944. doi: 10.1371/journal.pone.0086944

Schofield, B. R. (2011). "Central descending auditory pathways," in *Auditory and vestibular efferents*, (Springer), 261–290. doi: 10.1007/978-1-4419-7070-1_9

Sedley, W., Alter, K., Gander, P. E., Berger, J., and Griffiths, T. D. (2019). Exposing pathological sensory predictions in tinnitus using auditory intensity deviant evoked responses. *Journal of Neuroscience* 39, 10096–10103. doi: 10.1523/jneurosci.1308-19.2019

Sedley, W., Friston, K. J., Gander, P. E., Kumar, S., and Griffiths, T. D. (2016). An integrative tinnitus model based on sensory precision. *Trends in Neurosciences* 39, 799–812. doi: 10.1016/j.tins.2016.10.004

Serafini, R. A., Pryce, K. D., and Zachariou, V. (2020). The mesolimbic dopamine system in chronic pain and associated affective comorbidities. *Biological Psychiatry* 87, 64–73. doi: 10.1016/j.biopsych.2019.10.018

Sheppard, A., Stocking, C., Ralli, M., and Salvi, R. (2020). A review of auditory gain, low-level noise and sound therapy for tinnitus and hyperacusis. *International journal of audiology* 59, 5–15. doi: 10.1080/14992027.2019.1660812

Sherman, S. M. (2001). A wake-up call from the thalamus. *Nature neuroscience* 4, 344. doi: 10.1038/85973

Smith, P. H., Uhlrich, D. J., Manning, K. A., and Banks, M. I. (2012). Thalamocortical projections to rat auditory cortex from the ventral and dorsal divisions of the medial geniculate nucleus. *Journal of Comparative Neurology* 520, 34–51. doi: 10.1002/cne.22682

Song, J. J., Vanneste, S., and De Ridder, D. (2015). Dysfunctional noise cancelling of the rostral anterior cingulate cortex in tinnitus patients. *PloS one* 10,

Tan, C. M., Lecluyse, W., McFerran, D., and Meddis, R. (2013). Tinnitus and patterns of hearing loss. *J Assoc Res Otolaryngol* 14, 275–282. doi: 10.1007/s10162-013-0371-6

Tang, A. D., Lowe, A. S., Garrett, A. R., Woodward, R., Bennett, W., Canty, A. J., et al. (2016). Construction and evaluation of rodent-specific rTMS coils. *Frontiers in neural circuits* 10:47.

Thomas, M. E., Guercio, G. D., Drudik, K. M., and de Villers-Sidani, É (2019). Evidence of hyperacusis in adult rats following non-traumatic sound exposure. *Frontiers in systems neuroscience* 13:55.

Turner, J. G., Brozoski, T. J., Bauer, C. A., Parrish, J. L., Myers, K., Hughes, L. F., et al. (2006). Gap detection deficits in rats with tinnitus: a potential novel screening tool. *Behavioral neuroscience* 120, 188. doi: 10.1037/0735-7044.120.1.188

Uylings, H. B., and van Eden, C. G. (1991). "Qualitative and quantitative comparison of the prefrontal cortex in rat and in primates, including humans," in *Progress in brain research*, (Elsevier), 31–62. doi: 10.1016/s0079-6123(08)62675-8

van der Loo, E., Gais, S., Congedo, M., Vanneste, S., Plazier, M., Menovsky, T., et al. (2009). Tinnitus intensity dependent gamma oscillations of the contralateral auditory cortex. *PLoS One* 4:e7396. doi: 10.1371/journal.pone.0007396

Vanneste, S., Alsalman, O., and De Ridder, D. (2019). Top-down and bottom-up regulated auditory phantom perception. *Journal of Neuroscience* 39, 364–378. doi: 10.1523/jneurosci.0966-18.2018

Vanneste, S., Plazier, M., Van de Heyning, P., and De Ridder, D. (2011). Repetitive transcranial magnetic stimulation frequency dependent tinnitus improvement by double cone coil prefrontal stimulation. *Journal of Neurology, Neurosurgery & Psychiatry* 82, 1160–1164. doi: 10.1136/jnnp.2010.213959

Vogler, D. P., Robertson, D., and Mulders, W. H. (2011). Hyperactivity in the ventral cochlear nucleus after cochlear trauma. *Journal of Neuroscience* 31, 6639–6645. doi: 10.1523/jneurosci.6538-10.2011

Volz, L. J., Benali, A., Mix, A., Neubacher, U., and Funke, K. (2013). Dose-dependence of changes in cortical protein expression induced with repeated transcranial magnetic theta-burst stimulation in the rat. *Brain stimulation* 6, 598–606. doi: 10.1016/j.brs.2013.01.008

Wagner, T. A., Zahn, M., Grodzinsky, A. J., and Pascual-Leone, A. (2004). Three-dimensional head model simulation of transcranial magnetic stimulation. *IEEE Transactions on Biomedical Engineering* 51, 1586–1598. doi: 10.1109/tbme.2004.827925

Winkowski, D. E., Nagode, D. A., Donaldson, K. J., Yin, P., Shamma, S. A., Fritz, J. B., et al. (2018). Orbitofrontal cortex neurons respond to sound and activate primary auditory cortex neurons. *Cerebral Cortex* 28, 868–879. doi: 10.1093/cercor/bhw409

Wu, C., Martel, D. T., and Shore, S. E. (2016). Increased synchrony and bursting of dorsal cochlear nucleus fusiform cells correlate with tinnitus. *Journal of Neuroscience* 36, 2068–2073. doi: 10.1523/jneurosci.3960-15.2016

Yu, X.-J., Xu, X.-X., He, S., and He, J. (2009). Change detection by thalamic reticular neurons. *Nature neuroscience* 12, 1165. doi: 10.1038/nn.2373

Yushkevich, P. A., Gao, Y., and Gerig, G. (2016). "ITK-SNAP: An interactive tool for semi-automatic segmentation of multi-modality biomedical images," in *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, (IEEE).

Zikopoulos, B., and Barbas, H. (2006). Prefrontal projections to the thalamic reticular nucleus form a unique circuit for attentional mechanisms. *Journal of Neuroscience* 26, 7348–7361. doi: 10.1523/jneurosci.5511-05.2006

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Zimdahl, Thomas, Bolland, Leggett, Barry, Rodger and Mulders. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 26 July 2021
doi: 10.3389/fnins.2021.661477

Cross-Modal Interaction Between Auditory and Visual Input Impacts Memory Retrieval

Viorica Marian[1], Sayuri Hayakawa[1] and Scott R. Schroeder[1,2]*

[1] *Department of Communication Sciences and Disorders, Northwestern University, Evanston, IL, United States,*
[2] *Department of Speech-Language-Hearing Sciences, Hofstra University, Hempstead, NY, United States*

How we perceive and learn about our environment is influenced by our prior experiences and existing representations of the world. Top-down cognitive processes, such as attention and expectations, can alter how we process sensory stimuli, both within a modality (e.g., effects of auditory experience on auditory perception), as well as across modalities (e.g., effects of visual feedback on sound localization). Here, we demonstrate that experience with different types of auditory input (spoken words vs. environmental sounds) modulates how humans remember concurrently-presented visual objects. Participants viewed a series of line drawings (e.g., picture of a cat) displayed in one of four quadrants while listening to a word or sound that was congruent (e.g., "cat" or <meow>), incongruent (e.g., "motorcycle" or <vroom–vroom>), or neutral (e.g., a meaningless pseudoword or a tonal beep) relative to the picture. Following the encoding phase, participants were presented with the original drawings plus new drawings and asked to indicate whether each one was "old" or "new." If a drawing was designated as "old," participants then reported where it had been displayed. We find that words and sounds both elicit more accurate memory for *what* objects were previously seen, but only congruent environmental sounds enhance memory for *where* objects were positioned – this, despite the fact that the auditory stimuli were not meaningful spatial cues of the objects' locations on the screen. Given that during real-world listening conditions, environmental sounds, but not words, reliably originate from the location of their referents, listening to sounds may attune the visual dorsal pathway to facilitate attention and memory for objects' locations. We propose that audio-visual associations in the environment and in our previous experience jointly contribute to visual memory, strengthening visual memory through exposure to auditory input.

Keywords: multisensory integration, cross-modal interaction, audio-visual processing, auditory experience, visual memory, spatial memory, spoken words, environmental sounds

OPEN ACCESS

Edited by:
Miguel A. Merchán, University of Salamanca, Spain

Reviewed by:
Huizhong Whit Tao, University of Southern California, Los Angeles, United States
Chrysa Retsa, Centre Hospitalier Universitaire Vaudois (CHUV), Switzerland

****Correspondence:***
Sayuri Hayakawa
sayuri.hayakawa@northwestern.edu

Specialty section:
This article was submitted to Auditory Cognitive Neuroscience, a section of the journal Frontiers in Neuroscience

Received: *30 January 2021*
Accepted: *24 June 2021*
Published: *26 July 2021*

Citation:
Marian V, Hayakawa S and Schroeder SR (2021) Cross-Modal Interaction Between Auditory and Visual Input Impacts Memory Retrieval. Front. Neurosci. 15:661477. doi: 10.3389/fnins.2021.661477

INTRODUCTION

Many of us have had the experience of feeling transported back in time upon exposure to a familiar sight or sound – a song on the radio might conjure the image of your first car, or the sight of a tuba might invoke the cacophony of your middle school band. Such phenomena illustrate the essentially multisensory quality of what we experience, and subsequently, what we remember. While it is

simple enough to intuit that memories made in one sensory modality could become associated with those in another by virtue of their shared context or source, here we ask whether the things we hear can directly alter how we encode and remember the things that we see.

Contrary to the traditional view of sensory processing as largely modality-specific and "bottom-up" in nature (e.g., from a sense organ up through modality-specific subcortical and cortical areas), there is now considerable evidence that dynamic networks of descending and lateral pathways enable higher-level functions (e.g., attention) to influence and optimize even very basic sensory processes (e.g., cochlear function in animals, Maison and Liberman, 2000; Darrow et al., 2006; Delano et al., 2007, and humans, e.g., Marian et al., 2018b), as well as the integration of inputs across modalities (most extensively researched with auditory and visual stimuli; Stein et al., 1996; Giard and Peronnet, 1999; Calvert et al., 2000; McDonald et al., 2000; Calvert, 2001; Molholm et al., 2002, 2004; Schroeder and Foxe, 2005; Driver and Noesselt, 2008; Talsma et al., 2010; Moran et al., 2013; Ten Oever et al., 2016). This work has contributed to our current understanding of sensory perception as an on-going interplay between stimulus-driven processing and top-down influence, both of which are characterized by significant cross-modal interactivity. Relatively less is known, however, regarding the nature of cross-modal interactivity and the role of perceptual experience in memory. The present study was therefore designed to examine the joint impact of multisensory exposure and experience on visual and spatial memory.

Audio-Visual Interactions in Visual Perception

Robust behavioral evidence confirms that a range of visual processes, including detection, identification, and localization can be facilitated by the concurrent processing of auditory stimuli (Stein et al., 1996; Vroomen and De Gelder, 2000; Spivey et al., 2001; Molholm et al., 2004; Seitz et al., 2006; Van der Burg et al., 2008; Lupyan and Spivey, 2010; Salverda and Altmann, 2011; see also Shams et al., 2000; Morein-Zamir et al., 2003 for examples of visual distortion by auditory stimuli). There remains, however, considerable ambiguity and debate regarding the mechanisms underlying cross-modal facilitation, including the relative contributions of stimulus-driven bottom-up processing vs. top-down control (see De Meo et al., 2015 for discussion).

Cross-modal interactions of sensory information can occur at multiple stages of processing. Most commonly, we think of audio-visual integration as occurring in higher-level multisensory areas (e.g., superior temporal sulcus, middle temporal gyrus, inferior parietal cortex) where auditory and visual information about behaviorally relevant stimuli (e.g., objects and speech) can be brought together from separate processing streams to be integrated into a coherent percept (e.g., Calvert et al., 2000; Raij et al., 2000; Calvert, 2001; Beauchamp et al., 2004a,b; Powers et al., 2012). However, integration can also occur at lower levels, including in regions traditionally thought to be unisensory processing areas (e.g., the primary visual cortex V1; Giard and Peronnet, 1999; Molholm et al., 2002; Schroeder and Foxe, 2005; Watkins et al., 2006; Murray et al., 2016 for review), as well as in subcortical regions (e.g., multisensory neurons in the superior colliculus, Miller and D'Esposito, 2005), which receive descending projections from modality-specific subregions and can play a key role in stimulus localization and the control of orienting behaviors (Wallace, 2004). Furthermore, cross-modal interaction can be initiated at multiple stages concurrently in response to different factors (e.g., stimulus characteristics, contextual factors, top-down influences), and can shape perception and behavior via a variety of distinct mechanisms (e.g., attentional orienting, multisensory integration, cross-modal influence; see De Meo et al., 2015; Ten Oever et al., 2016).

The mechanisms behind multisensory enhancement can be particularly ambiguous when either of the cross-modal inputs (on its own) would be sufficient to elicit a correct response (e.g., about an object's location or identity; see Driver and Noesselt, 2008). For instance, it has been shown that people are better at identifying what object they are seeing (e.g., a cat) if they are provided with a redundant auditory cue (e.g., the sound of a cat, <meow>; Chen and Spence, 2010). One explanation for this type of behavioral enhancement is that cross-modal interactions occur at early stages of processing, whereby exposure to the auditory stimulus automatically boosts attentional and perceptual processing of the visual input, strengthening the visual representation and facilitating identification. Alternatively, each signal could be processed up to and converge at the level of the stimulus meaning (visual image of a cat → visual representation of a cat → semantic concept of a cat ← auditory representation of <meow> ← auditory <meow>), and the dual activation of the semantic concept can make it easier to identify the image – either by engaging top-down processes to directly modulate perception or by affecting the decision or response criteria (without further involvement of perceptual/attentional processes; see **Figure 1**). Indeed, assuming the auditory input is a consistently reliable cue, it would be possible to correctly identify the "visual" target even if someone were to close their eyes, as the auditory stimulus would provide equally valid information about the object's identity. In other words, while redundant signals could involve a high degree of cross-modal integration and feedback to affect modality-specific processing (e.g., boosting processing of the visual features), it could also be primarily bottom-up in nature and would not even necessarily require the integration of cross-modal information.

Now consider a case in which a person is asked to identify an object's *location* from an array of pictures and is provided with the target ahead of time via a written word (e.g., cat). Assuming that was the only information the person received, the target word would be processed first (e.g., written word *cat* → orthographic representation of word *cat* → semantic concept of cat → visual features of a cat), which could then prepare the visual system to look for the corresponding visual object (e.g., by increasing sensitivity to the visual representation). In addition to top-down facilitation resulting from a previously presented target cue, visual search can be speeded if a congruent auditory sound (e.g., <meow>, or "cat") is presented along with

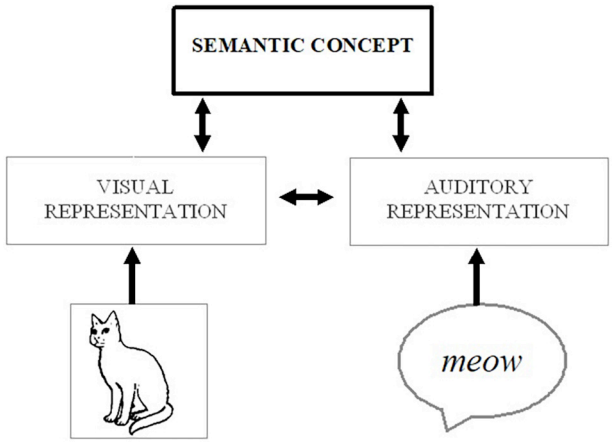

FIGURE 1 | Possible processing route whereby activation from auditory (e.g., a meowing sound) or visual (e.g., a picture of a cat) stimuli could spread to a corresponding conceptual representation (e.g., of a cat) and associated visual features, thereby strengthening the salience of the visual object.

the visual display, even if it provides no information about the object's location (Iordanescu et al., 2010; Lupyan and Spivey, 2010; Iordanescu et al., 2011). As with object identification, the auditory stimulus will begin with feed-forward activation, likely including associated visual features (e.g., auditory sound <meow> → auditory representation <meow> → concept of a cat → visual representation of a cat), which combined with the visual representation activated by the initial target word and the visual features in the display, can further boost activation and salience of the target object.

The key difference between these examples of object identification and localization is that in the latter case, the auditory cue is not sufficient to complete the task without any interaction with visual processing. In order to facilitate object localization, the auditory cue must improve how efficiently the *visual system* can take the final step of identifying the target's location. In other words, because the auditory input does not itself provide information about the task-relevant external property (i.e., the location), we can conclude that if it facilitates localization, it must be doing so by directly modulating how the visual information is processed. Indeed, there is evidence that this kind of cross-modal interaction can occur at even more basic levels of processing, such as speeded detection of a visual target following a color change and a concurrent spatially uninformative sound (e.g., a beep, or "pip"; Van der Burg et al., 2008). In this case, the temporal congruence between the sound and a color change can boost visual attention at the critical time, helping to identify the location of the visual target. A burst of noise can also increase the perceived intensity of a concurrently-presented light in a different location (Stein et al., 1996), showing that auditory inputs can modulate the perceptual quality of visual stimuli. Lastly, hearing a sound can not only facilitate detection of a visual target appearing in the same place at the same time (Driver and Spence, 1998; Bolognini et al., 2005), but also when a visual target appears in a location *previously* cued by a sound (McDonald et al., 2000), providing strong evidence that auditory inputs can elicit top-down attentional orienting to a specific location, subsequently enhancing the perceptual salience of spatially coincident visual inputs.

Audio-Visual Interactions in Visual Memory

While there is now substantial behavioral and neural evidence that cross-modal inputs can directly modulate attentional and perceptual processes, including perceptual learning of simple, artificial stimuli (e.g., dots and beeps; see Shams and Seitz, 2008 for review), relatively less is known about the influence of multisensory integration on memory for semantically meaningful, naturalistic objects. Given that multisensory integration has been shown to enhance the salience and attentional processing of sensory stimuli (Parkhurst et al., 2002; Santangelo and Spence, 2007; Matusz and Eimer, 2011; see Koelewijn et al., 2010; Talsma et al., 2010 for reviews), which can, in turn, strengthen memory encoding (Salthouse et al., 1984; Evans and Baddeley, 2018), it would be reasonable to expect that auditory facilitation of visual attention and perception could translate to better visual memory as well. Indeed, audio-visual encoding can enhance recognition memory for visual objects (Murray et al., 2004, 2005; Lehmann and Murray, 2005; Thelen et al., 2012, 2014, 2015; Moran et al., 2013; Thelen and Murray, 2013; Heikkilä et al., 2015; Matusz et al., 2015; Ueno et al., 2015; see Matusz et al., 2017 for review). For instance, in a series of experiments utilizing a continuous recognition task (identifying "old" vs. "new" pictures), Matusz et al. (2017) found that unimodal pictures (e.g., a cow) that were initially encoded along with a task irrelevant, but semantically congruent characteristic sound (e.g., "moo") were later recognized with greater accuracy than unimodal stimuli or images paired with incongruent (e.g., "meow") or neutral sounds (e.g., a tone).

Meyerhoff and Huff (2016) similarly found that recognition of semantically congruent audio-visual film clips was greater than of incongruent audio-visual clips. Unlike Matusz et al. (2017), however, memory for both congruent and incongruent audio-visual clips exceeded that of unimodal clips. The authors speculate that the relatively greater advantage for congruent audio-visual stimuli may stem from the multisensory integration of auditory and visual information in memory, resulting in more elaborate representations (analogous to perceptual advantages observed for congruent audio-visual stimuli; e.g., Chen and Spence, 2010). However, unlike perceptual integration, which typically requires precise temporal correspondence, audio-visual facilitation in memory was found to persist despite temporal asynchrony, leading the authors to conclude that cross-modal effects on memory are unlikely to be mere extensions of perceptual processes.

Advantages of multisensory encoding have also been extended to motion. Recognition memory is superior for moving relative to static images (Matthews et al., 2007). Similar to Matthews et al. (2007), Meyerhoff and Huff (2016) propose that the superior memory for moving pictures may result from the construction and storage of a scene-based "object file" (Hollingworth and

Henderson, 2002) in long-term memory, which contains detailed representations of not only the visual forms of particular objects, but also their spatial positions within a larger scene. Hollingworth and Henderson (2002) theorized that visual fixations to different components of a scene play a key role in the formation of object files and that directing visual fixations and attention to the spatial locations of previously seen objects can facilitate visual memory retrieval. Matthews et al. (2007) build upon this model by proposing that visual objects in a scene are encoded along with not only spatial information, but temporal information as well, and that the activation of "motion schemata" facilitates subsequent recall of associated visuospatial memory traces.

These findings demonstrate that, like perceptual and attentional processes, memory can be influenced by exposure to multisensory stimuli. Unlike effects on perception, however, which have been documented using a wide variety of tasks (e.g., discrimination, localization, and detection), much of the existing work on multisensory memory has been limited to object recognition. Analogous to perceptual effects emerging from exposure to redundant audio-visual cues, enhanced visual recognition following congruent audio-visual encoding could result from the availability of two valid sources of information regarding an object's identity (i.e., the visual and auditory memory trace) rather than better memory of the visual percept itself. In addition to potentially providing additional retrieval cues (if both auditory and visual inputs are presented at test), exposure to one cue during retrieval could initiate the rapid reactivation of the other, both of which could be used to recall the identity of a previously seen object (e.g., the *redintegration hypothesis of memory retrieval*; Nyberg et al., 2000; Wheeler et al., 2000; Moran et al., 2013).

What has yet to be determined is whether audio-visual interactions and top-down attentional allocation to visual inputs at the encoding stage could facilitate memory for an object's features and visual context – in other words, whether hearing an auditory stimulus can influence how well the unimodal visual representation itself is remembered. The present study investigates this possibility by examining whether hearing a spatially uninformative auditory cue (i.e., a sound that does not correspond to a visual object's position on the screen) can improve visual memory for an object's *location*. If, as predicted, memory for an object's location is enhanced by listening to a corresponding sound, this facilitation is unlikely to result from the rapid retrieval of the spatially invalid auditory memory trace. Instead, it would suggest that the visual memory trace itself is strengthened by audio-visual encoding, providing compelling evidence of cross-modal interactivity in sensory memory.

The Role of Experience in Audio-Visual Interactions

In addition to examining how cross-modal inputs interact in memory, a second goal of the present research is to determine whether the impact of auditory stimuli on visual memory varies as a function of prior experience with particular types of sounds.

Audio-visual integration has long been known to be moderated by physical properties of stimuli in the environment, most notably spatial and temporal contiguity (with greater integration for inputs that are closer in space and time; Meredith and Stein, 1986; Frens et al., 1995; Royal et al., 2009), as well as by characteristics such as the stimuli's motion relative to the observer (e.g., greater integration for those that are looming than receding; Cappe et al., 2009). These principles are ecologically sensible given that real-world stimuli originating from the same source are likely to correspond spatially and temporally, and approaching stimuli can pose a potential danger (e.g., a predator or a car), making efficient processing especially consequential. Similarly, the semantic meanings attached to sensory stimuli in real-world contexts can provide information about the probability that multiple inputs represent features of the same object (e.g., the sight and sound of a firetruck), and may be especially likely to reveal experience-dependent differences in cross-modal interactivity (e.g., between particular types of auditory and visual input). Due to repeated experience associating visual and auditory features of objects (e.g., seeing a cat while hearing "meow"), object-based processing of auditory stimuli may boost activation of the corresponding visual representation, thereby increasing its salience through top-down and/or lateral feedback mechanisms (see Iordanescu et al., 2010, 2011). This experience-based explanation is consistent with Iordanescu et al.'s (2011) finding that cross-modal facilitation occurs between commonly co-occurring forms of object-based stimuli (e.g., visual features and sounds of objects as we interact with them, visual features and vocalized labels, such as when naming objects; and written and vocalized labels, such as when reading aloud), but not between stimuli that are not commonly processed together during real-world experiences (written labels and characteristic sounds). In other words, audio-visual interactivity during sensory processing varies as a function of our prior experiences with specific combinations of auditory and visual stimuli.

Prior experience with particular forms of auditory input (e.g., characteristic sounds vs. linguistic labels) can additionally modulate the types of representations that are brought to mind, which can subsequently impact performance on visual tasks. For instance, Lupyan and Thompson-Schill (2012) propose that the concepts activated by words tend to be more categorical and prototypical than those activated by characteristic sounds. In a series of studies, the authors observed that when participants were instructed to indicate whether a picture was congruent or incongruent with an auditory cue, performance was enhanced when the visual objects were cued by a verbal label (e.g., "cat") relative to a characteristic sound (e.g., a meowing sound; Experiments 1A–C). Importantly, this advantage for spoken words was greater for pictures that were rated as more "typical" representations of their referents, providing support for the hypothesis that words tend to activate prototypical exemplars. Furthermore, it was found that the label advantage was specific to nouns (e.g., the word "cat"), and was not found for verbs (e.g., the word "meowing") or verbalized sound imitations (e.g., the word "meow"; Experiment 2). Together, these findings provide support for the notion that linguistic labels are distinct from characteristic sounds in that they are more likely to represent an abstract concept that encompasses any number of individual exemplars (e.g., the general concept of a dog), whereas a characteristic

sound (e.g., of barking) is likely to invoke a more specific referent (e.g., a particular type of dog; see Waxman and Gelman, 2009). As a result, the concepts that are brought to mind in response to a linguistic label should be less idiosyncratic relative to characteristic sounds, which may subsequently facilitate the initial recognition of any given visual depiction.

The goal of the present study is twofold. First, we examine whether cross-modal facilitation observed during perceptual and attentional tasks extends to subsequent memory. Second, we examine whether effects of auditory input on visual memory vary as a function of prior experience with particular types of sounds. Specifically, we examine the possibility that the more concrete and exemplar-specific nature of characteristic sounds may, in some cases, promote *better* memory compared to linguistic labels, such as when attempting to remember where objects were previously seen. According to Edmiston and Lupyan (2015), environmental sounds can be considered "motivated" cues, in that the qualities of the auditory stimuli convey meaningful information regarding the physical source, including where it is relative to the observer. Spoken words, in contrast, are "unmotivated" cues in that they do not provide information about the specific physical source – while there are certainly situations in which one hears an object's label while looking at its referent (e.g., "that's my cat"), the referent is rarely the *source* of spoken words and is very often entirely absent. Having learned over years of experience that one is likely to see the physical features associated with an auditory stimulus (e.g., a cat) upon orienting to the location of an environmental sound (e.g., <meow>), but not a word (e.g., "cat"), it is possible that sounds will be more effective at engaging attentional processes dedicated to visuospatial localization compared to words. We therefore investigate whether memory for the locations of visual objects may be greater when they are initially encoded along with an environmental sound compared to a spoken word, even when both auditory cues are spatially uninformative.

The Present Study

The present study was designed to test the following hypotheses regarding the nature of audio-visual interactivity in visual object memory:

Hypothesis 1: Semantic congruence between auditory and visual inputs will facilitate visual memory.

Specifically, in addition to facilitating recognition (i.e., "what") of visual objects (replicating earlier findings), we predict that spatially uninformative auditory input will improve memory for objects' locations, providing evidence that auditory input can modulate visual memory in the absence of redundant cues.

Hypothesis 2: Cross-modal facilitation will vary as a function of experience-dependent associations between particular auditory inputs (spoken words vs. environmental sounds) and specific visuospatial dimensions ("what" vs. "where").

While recognition memory is expected to be facilitated by both spoken words and environmental sounds (due to their informational relevance for identifying *what* objects were seen), spatial memory may be selectively enhanced by environmental sounds due to repeated experience associating the spatial locations of visual objects with environmental sounds, but not words.

MATERIALS AND METHODS

Participants

Forty-three young adults (mean age = 21.9; *SD* = 3.2; 79% female) participated in the experiment[1]. Memory for visual objects associated with spoken words and environmental sounds were completed in two separate blocks, with the order of the blocks counterbalanced across subjects. All participants provided written consent and the research reported in this manuscript was approved by the University Institutional Review Board (STU00023477).

Stimuli

Different sets of stimuli were used for the congruent, incongruent, and neutral trials of the two experimental blocks. The spoken word block included two types of neutral stimuli (a tone and a pseudoword), while the environmental sound block included one type of neutral stimulus (a tone). The inclusion of both pseudowords and tones as baselines in the spoken word condition enabled us to examine potential differences between neutral verbal and non-verbal cues within the same experimental context, while also allowing us to directly compare the effects of meaningful spoken words vs. environmental sounds relative to the same baseline (i.e., neutral tones). The procedures were identical for spoken words and environmental sounds.

During the encoding phase, participants were presented with either 64 (for spoken words) or 60 (for environmental sounds) black and white pictures of objects selected from the *International Picture Naming Project Database* (Szekely et al., 2004). All pictures were similar in saturation and line thickness and were displayed on a computer screen with 2,650 × 1,440 resolution, with participants seated 80 cm away from the screen. Labels representing each of the objects in the spoken word and environmental sound blocks were matched on English frequency (SUBTLEXUS; Brysbaert and New, 2009), concreteness, familiarity, and imageability (MRC Psycholinguistic Database; Coltheart, 1981) across the spoken word and environmental sound blocks, as well as across lists within each block (see **Supplementary Tables 1–3** for details). Picture-word and picture-sound pairs spanned many different categories (e.g., animals, instruments, and food) and care was taken to ensure that there was a similar number of items per semantic category across lists within each block[2].

Spoken words had a mean duration of 801.99 ms (*SD* = 134.04), ranged from 445.99 ms to 997.89 ms, and were recorded at 44,100 Hz by a female native English speaker.

[1]Data from two participants who were included in environmental sound analyses were not analyzed for spoken words due to technical errors.

[2]The number of items per category varied across blocks due to the constraint that objects in the environmental sound category needed to be associated with a recognizable sound. Follow-up analyses were therefore conducted on a subset of items that were matched in the number of items per category across the word and sound blocks (see **Supplementary Table 4** for additional detail).

Environmental sounds had a duration of 1,000 ms. Neutral tones were 1,000 ms sine waveforms ranging from 250 to 1,750 Hz in 100 Hz increments in the spoken word condition and from 300 to 2,200 Hz in 100 Hz increments in the environmental sound condition. Different tones were utilized on each trial in order to mimic the structure of the congruent, incongruent, and neutral spoken word trials and the congruent and incongruent environmental sound trials, where auditory cues varied from trial to trial and between the spoken word and environmental sound conditions. In this way, each tone, word, or sound was paired with a single picture regardless of condition or trial type. Every trial additionally included audio-visual stimuli during encoding, but only unimodal visual stimuli during retrieval (see "Procedure" for additional detail regarding the retrieval phase). Prior research has shown that memory is enhanced when the context of retrieval matches that of encoding (i.e., context-dependent memory; see Smith and Vela, 2001 for review), as well as for odd events and items that are distinct with respect to the surrounding context or stimuli (i.e., a distinctiveness or isolation effect; von Restorff, 1933; Dunlosky et al., 2000; Hunt and Lamb, 2001). To ensure that the sensory contexts of encoding were equally probable and dissimilar to those of retrieval across the different trial types, neutral words and sounds were chosen as controls in lieu of unimodal visual stimuli (the latter of which could benefit from a match in encoding and retrieval contexts and their relative distinctiveness during encoding). All auditory cues were amplitude normalized and presented through headphones using monophonic sound reproduction, so as to make them spatially uninformative. In other words, while the auditory cues did contain spatial information and would be perceived to emanate from the center point of the display, the spatial location of the sound was fixed across all trials and did not provide meaningful information regarding the spatial location of the visual object on the screen. The sound level was fixed at two bars on an iMac desktop computer for all participants.

The visual display was divided into 9 equally sized grid spaces, following a 3 × 3 grid. On each trial of both word and sound blocks, a single picture was presented in one of four positions (top left corner, top right corner, bottom left corner, bottom right corner, with each critical position separated by an empty grid space). The location of pictures was randomized across trials with the constraint that pictures on consecutive trials never appeared in the same spatial location.

Spoken Words

Each picture in the spoken word block was presented concurrently with an auditory cue in one of four trial types (16 trials each): *congruent word* (i.e., the English label for the depicted object; e.g., a picture of a shoe + "shoe"), *incongruent word* (i.e., the English label for an object from a different semantic category; e.g., a picture of a guitar + "apple"; see **Supplementary Table 4**), *neutral word* (i.e., a meaningless pseudoword; e.g., a picture of a snake + "fenip"), or *neutral tone* (i.e., a meaningless beep; e.g., a picture of a candle + a tonal beep).

Five lists of 16 objects and one list of pseudowords were compiled to create the picture-auditory cue pairs (see **Table 1**). Pseudowords were taken from Bartolotti and Marian (2012), were constructed to follow English phonotactic rules and matched in length ($M = 5.94$ letters) to the real word stimuli used in the present experiment ($M = 5.88$ letters; $p = 0.902$). Each of the five object lists was used as the visual or auditory stimulus set in one of the trial types. To illustrate, a participant may see pictures from List 1 paired with words from List 1 (*congruent word*), pictures from List 2 paired with words from List 3 (*incongruent word*), pictures from List 4 paired with pseudowords (*neutral word*), and pictures from List 5 paired with a tonal beep (*neutral tone*). The lists were rotated across participants; each of the lists served in each position an equal number of times across participants. The 64 trials were presented in 16 four-trial runs. Each of the four trial types was presented in each run, with the order of trial types counterbalanced across runs (i.e., each run included a congruent word, incongruent word, neutral word, and neutral tone trial in a different order).

Environmental Sounds

Each picture in the environmental sound block was presented concurrently with an auditory cue in one of three trial types (20 trials each): *congruent sound* (i.e., an environmental sound corresponding to the depicted object; e.g., a picture of a dog + a sound made by a dog <woof–woof>), *incongruent sound* (i.e., an environmental sound corresponding to an object from a different semantic category; e.g., a picture of a trumpet + a sound made by a motorcycle <vroom–vroom>), or *neutral sound* (i.e., a meaningless beep; e.g., a picture of a helicopter + a tonal beep).

Four lists of 20 objects (unique from those used for spoken words) were compiled to create the picture-sound pairs (see **Table 2**). Each of the four lists was rotated across participants to serve as the visual or auditory stimulus set in one of the trial types. The 60 trials were presented in 20 three-trial runs, with each run including each of the three trial types in a counterbalanced order (congruent sound, incongruent sound, neutral sound).

TABLE 1 | List of the spoken word stimuli used in the present study.

List 1	List 2	List 3	List 4	List 5	Pseudowords
Apple	Belt	Bride	Anchor	Bra	Acrip
Cast	Book	Church	Button	Branch	Appint
Castle	Cowboy	Desk	Cloud	Chair	Bakloo
Cigarette	Finger	Giraffe	Doctor	Chimney	Eazoond
Dinosaur	Flag	Glue	Funnel	Dentist	Fenip
Doll	Grapes	Hat	Hammock	Diaper	Fummawp
Dress	Hamburger	Mask	Igloo	Eggplant	Ganteh
Ear	Leaf	Mountain	Kangaroo	Lamp	Glolay
Eel	Lemon	Mushroom	King	Llama	Iyork
Magnet	Medal	Orange	Puzzle	Map	Lateep
Pear	Microscope	Pipe	Shoe	Needle	Munbo
Pillow	Mop	Pirate	Thermos	Nun	Nepri
Sandwich	Penguin	Pizza	Tire	Plate	Peftoo
Stethoscope	Pyramid	Porcupine	Tomato	Present	Shundoe
Submarine	Skeleton	Refrigerator	Watermelon	Shrimp	Toymeen
Unicorn	Tractor	Vest	Wig	Spaghetti	Unyops

TABLE 2 | List of the environmental sound stimuli used in the present study.

List 1	List 2	List 3	List 4
Airplane	Astronaut	Accordion	Ambulance
Basketball	Bomb	Banjo	Baby
Bell	Can	Cannon	Bicycle
Boat	Elephant	Dog	Chainsaw
Camera	Gorilla	Dragon	Chicken
Cat	Harmonica	Drill	Clock
Donkey	Horse	Glasses	Cow
Door	Lawn mower	Harp	Cymbals
Drums	Lips	Heart	Dolphin
Frog	Matches	Lightning	Duck
Goat	Monkey	Lion	Flute
Gun	Motorcycle	Microphone	Guitar
Hippopotamus	Owl	Microwave	Hammer
Jackhammer	Piano	Pig	Hands
Nose	Radio	Shower	Helicopter
Printer	Rain	Snake	Rocket
Robot	Telephone	Stapler	Seal
Saxophone	Tornado	Swords	Taco
Toilet	Trumpet	Train	Violin
Xylophone	Typewriter	Whistle	Whip

Procedure

Prior to beginning the task, participants were asked to remember the pictures for a later memory test while ignoring the auditory cues. At the beginning of each trial, a central fixation cross was presented for 200 ms, followed by the simultaneous presentation of an auditory cue and a picture, which remained on screen for 1,000 ms (see **Figure 2**).

After all encoding trials, participants completed a simple 5-min numerical filler task (determining which of two values is larger) before completing the retrieval task.

For the retrieval task, participants were presented with the 64 (spoken word) or 60 (environmental sound) pictures that appeared during the encoding task (i.e., "old" pictures), as well as an equal number of foil pictures that were not previously seen (i.e., "new" pictures). During the recognition phase of each trial, a single picture appeared in the center of the screen and participants were asked to click on *old* if the picture was previously seen during the encoding phase, and to click on *new* if it was not. If an image was designated as *old,* participants were asked to indicate which spatial location the picture appeared in during the encoding phase by clicking on one of four boxes located in the four corners of the screen (see **Figure 3**). Participants were instructed to respond as quickly as possible without sacrificing accuracy. Across the two blocks, participants completed a total of 124 encoding trials and 248 retrieval trials and the entire experiment lasted approximately 30 min.

Data Analysis

We began with separate analyses of spoken words and environmental sounds, followed by a combined analysis. For both words and sounds, two separate generalized linear mixed-effects models were constructed to examine the effects of trial type (congruent, incongruent, neutral) on (1) recognition ("what") and (2) location ("where") accuracy for pictures that were previously seen during the encoding phase. Trial type was entered as a fixed effect and treatment coded to compare each level to congruent trials (i.e., congruent [0] vs. incongruent [1], neutral word [1], and neutral tone [1] trials for spoken words and congruent [0] vs. incongruent [1] and neutral tone [1] trials for environmental sounds). Models additionally included random intercepts for participant and target, as well as word frequency (zipf), concreteness, imageability, and familiarity of the targets' labels as covariates.[3] Following initial analyses comparing congruent trials to incongruent and neutral trials, planned pairwise comparisons were conducted to compare each of the incongruent and neutral trial types to each other. Follow-up analyses were additionally conducted on a subset of items ($N = 35$ out of 80), which were matched in semantic category across the spoken word and environmental sound lists (see **Supplementary Table 4**).

RESULTS

Spoken Words

Recognition ("What")

Recognition accuracy was significantly higher on congruent trials relative to neutral tone trials (*Estimate* $= -0.33$, $SE = 0.14$, $z = -2.41$, $p = 0.016$) and marginally higher than neutral word trials (*Estimate* $= -0.27$, $SE = 0.14$, $z = -1.94$, $p = 0.052$; see **Figure 4**). Similarly, accuracy on incongruent trials was significantly higher than on neutral tone trials (*Estimate* $= -0.30$, $SE = 0.14$, $z = -2.18$, $p = 0.029$) and marginally higher than on neutral word trials (*Estimate* $= -0.25$, $SE = 0.14$, $z = -1.77$, $p = 0.076$). Accuracy did not differ between congruent and incongruent trials (*Estimate* $= -0.4$, $SE = 0.14$, $z = -0.29$, $p = 0.769$) or between neutral word and neutral tone trials (*Estimate* $= -0.08$, $SE = 0.14$, $z = -0.55$, $p = 0.582$)[4]. These findings indicate that memory for previously-seen objects is enhanced when they are paired with meaningful words, regardless of whether or not the words are congruent with the visual object. Concurrent presentation of meaningless non-words, on the other hand,

[3]Models including covariates plus random slopes for trial type failed to converge. Similar patterns of results were obtained when models excluded covariates and included random intercepts for participant and target, as well as by-participant and by-target random slopes for trial type (i.e., the maximal random effects structure, Barr et al., 2013).

[4]Similar patterns were observed when recognition accuracy scores for previously seen pictures were replaced with d'prime scores which additionally accounted for false alarms and correct rejections of new items. D'prime scores were higher for congruent trials relative to both neutral word [*Estimate* $= -0.14$, $SE = 0.07$, $t(120) = -2.05$, $p = 0.042$] and neutral tone trials [*Estimate* $= -0.19$, $SE = 0.07$, $t(120) = -2.82$, $p = 0.006$]. Likewise, d'prime scores were higher for incongruent trials relative to neutral word [*Estimate* $= -0.12$, $SE = 0.06$, $t(40) = -2.09$, $p = 0.043$] and neutral tone trials [*Estimate* $= -0.17$, $SE = 0.07$, $t(40) = -2.53$, $p = 0.016$]. Scores did not differ between congruent and incongruent trials or between neutral word and neutral tone trials ($ps > 0.05$).

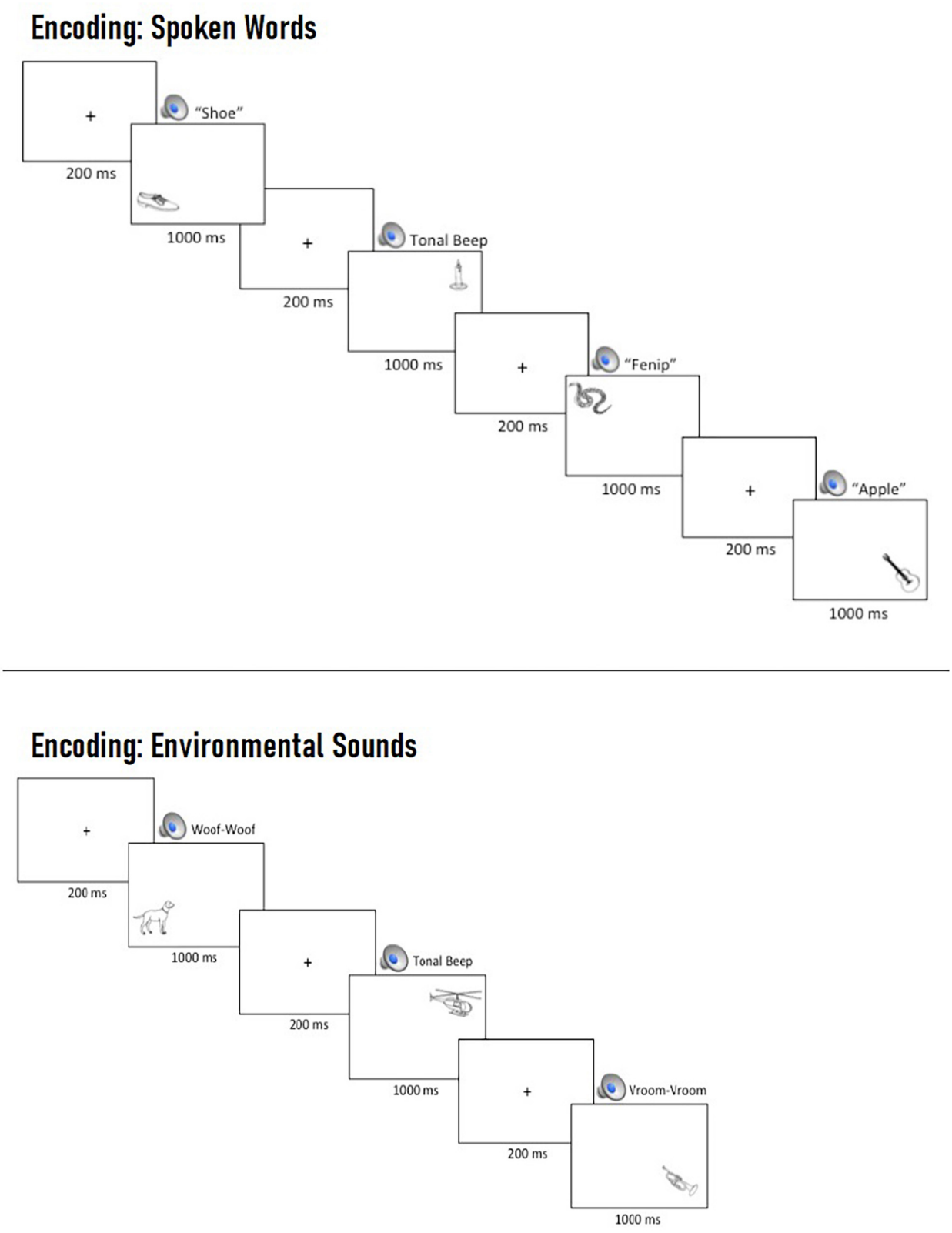

FIGURE 2 | Example multi-trial run of spoken words **(Top)** and environmental sounds **(Bottom)** during encoding. On each trial, participants were presented with a central fixation cross for 200 ms, which was replaced by the concurrent presentation of a task-irrelevant, spatially uninformative auditory cue and a picture in one of four locations, which remained on screen for 1,000 ms prior to the beginning of the next trial.

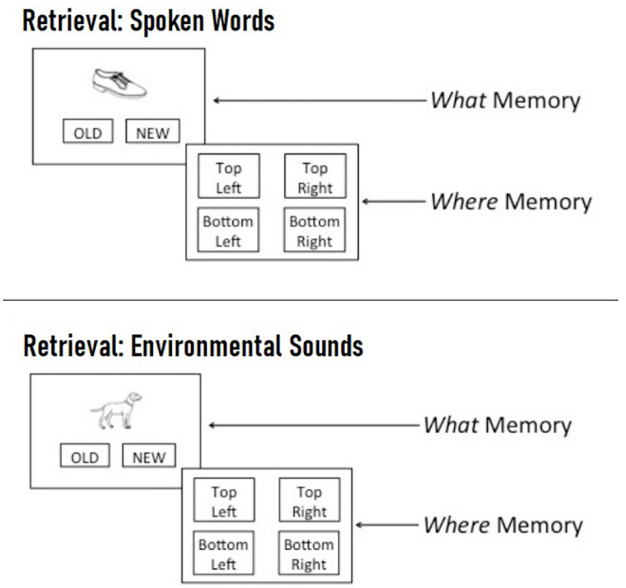

FIGURE 3 | Example spoken word **(Top)** and environmental sound **(Bottom)** retrieval trial. On each trial, participants were first presented with a single picture in the center of the screen and asked to indicate whether they recognized it from the encoding phase by clicking "old" or whether it was not previously seen by clicking "new." If a picture was designated as "old," participants were then asked to indicate the spatial location where the picture appeared during the encoding phase by clicking on one of four quadrants labeled "top left," "top right," "bottom left," or "bottom right."

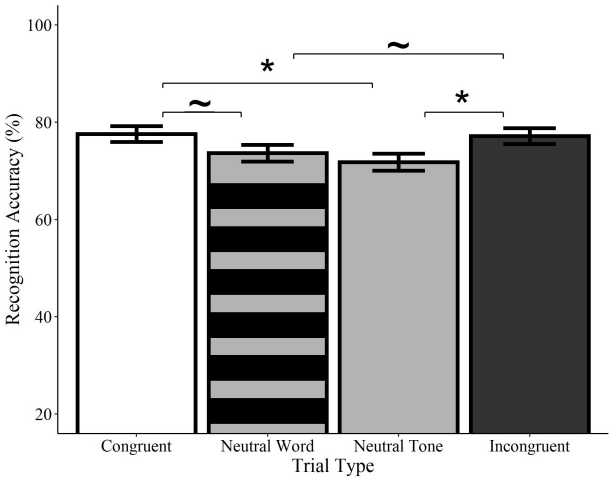

FIGURE 4 | Effect of trial type on recognition accuracy for spoken words. Visual objects initially paired with congruent (white) or incongruent (black) spoken words were recognized with significantly greater accuracy than those paired with neutral tones (solid gray) and with marginally greater accuracy than pictures paired with neutral words (striped gray). Accuracy did not differ between neutral tone and neutral word trials or between congruent and incongruent trials. Error bars represent standard error. $^{\sim}p < 0.10$, $^{*}p < 0.05$.

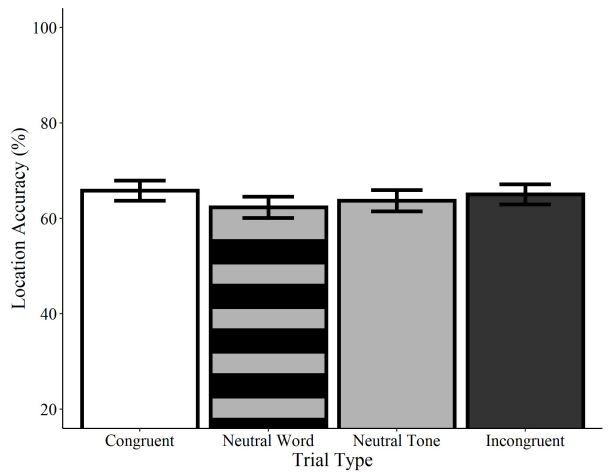

FIGURE 5 | Effect of trial type on location accuracy for spoken words. Location accuracy did not differ between congruent (white), neutral word (striped gray), neutral tone (solid gray), and incongruent (black) spoken word trials. Error bars represent standard error.

resulted in accuracy scores that were numerically lower than those of meaningful words, but higher than meaningless tones.

Follow-up analyses on category-matched items revealed a similar pattern, with significantly higher recognition accuracy on congruent trials relative to both neutral tone (*Estimate* = −0.53, *SE* = 0.21, $z = -2.48$, $p = 0.013$) and neutral word trials (*Estimate* = −0.47, *SE* = 0.21, $z = -2.19$, $p = 0.028$) and marginally higher accuracy on incongruent trials relative to neutral tone (*Estimate* = −0.40, *SE* = 0.20, $z = -1.93$, $p = 0.054$) and neutral word trials (*Estimate* = −0.36, *SE* = 0.21, $z = -1.71$, $p = 0.087$). Congruent trials did not differ from incongruent trials (*Estimate* = −0.09, *SE* = 0.21, $z = -0.40$, $p = 0.689$) and neutral tone trials did not differ from neutral word trials (*Estimate* = −0.08, *SE* = 0.20, $z = -0.42$, $p = 0.672$).

Location ("Where")

Analyses of location accuracy revealed no significant differences between congruent trials and incongruent word, neutral word, or neutral tone trials (*ps* > 0.05; see **Figure 5**). Likewise, no differences were observed between incongruent trials and neutral word and tone trials or between neutral word and neutral tone trials (*ps* > 0.05). Similarly, no effects of trial type were found in any comparisons using the category-matched subset of items (*ps* > 0.05). Contrary to the effects observed for recognition memory, these results indicate that accuracy for the locations of visual objects is not influenced by the concurrent presentation of congruent or incongruent words relative to neutral words or tones.

Environmental Sounds

Recognition ("What")

Recognition accuracy was significantly higher on congruent trials relative to neutral tone trials (*Estimate* = −0.41, *SE* = 0.12, $z = -3.34$, $p < 0.001$), as well as on incongruent trials relative to neutral tone trials (*Estimate* = −0.31, *SE* = 0.12, $z = -2.46$,

$p = 0.014$; see **Figure 6**). Accuracy did not differ between congruent and incongruent trials (*Estimate* $= -0.13$, $SE = 0.13$, $z = -0.99$, $p = 0.322$)[5]. Consistent with the analyses of spoken words, these findings indicate that recognition for *what* objects were previously seen is enhanced by the concurrent presentation of meaningful sounds, regardless of whether they are semantically congruent with the visual object.

Follow-up analyses on category-matched items revealed a similar pattern, with significantly higher recognition accuracy on congruent trials relative to neutral tone trials (*Estimate* $= -0.56$, $SE = 0.18$, $z = -3.06$, $p = 0.002$) and on incongruent trials relative to neutral tone trials (*Estimate* $= -0.45$, $SE = 0.19$, z = 02.45, $p = 0.014$). Congruent trials did not differ from incongruent trials (*Estimate* $= -0.12$, $SE = 0.19$, $z = -0.63$, $p = 0.530$).

Location ("Where")

Location accuracy was significantly higher on congruent trials relative to both incongruent (*Estimate* $= -0.54$, $SE = 0.12$, $z = -4.33$, $p < 0.0001$) and neutral tone trials (*Estimate* $= -0.58$, $SE = 0.13$, $z = -4.62$, $p < 0.0001$; see **Figure 7**). Incongruent trials did not differ from neutral tone trials (*Estimate* $= -0.04$, $SE = 0.12$, $z = -0.36$, $p = 0.722$). Similarly, analyses of category-matched items revealed significantly higher location accuracy for congruent trials relative to incongruent (*Estimate* $= -0.57$, $SE = 0.19$, $z = -.96$, $p = 0.003$) and neutral tone trials (*Estimate* $= -0.76$, $SE = 0.20$, $z = -3.85$, $p < 0.001$) and no difference between incongruent and neutral tone trials (*Estimate* $= -0.18$, $SE = 0.19$, $z = -0.93$, $p = 0.350$). In other words, memory for objects' locations is enhanced when they are initially encoded alongside a congruent, but not a neutral or incongruent sound, despite the fact that the sounds were not meaningful spatial cues.

[5]Similar patterns were observed for *d'* prime scores, which were significantly higher for congruent trials relative to neutral tone trials [*Estimate* $= -0.21$, $SE = 0.07$, $t(84) = -2.98$, $p = 0.004$], as well as for incongruent trials relative to neutral tone trials [*Estimate* $= -0.17$, $SE = 07$, $t(42) = -2.40$, $p = 0.021$]. Scores did not differ between congruent and incongruent trials ($p = 0.518$).

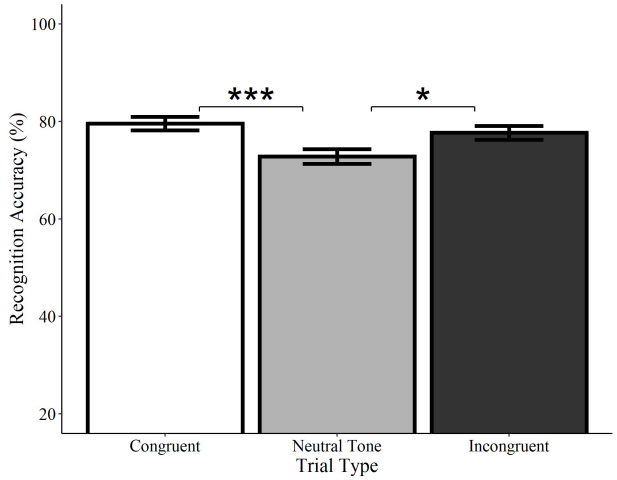

FIGURE 6 | Effect of trial type on recognition accuracy for environmental sounds. Visual objects initially paired with congruent (white) or incongruent (black) environmental sounds were recognized with significantly greater accuracy than those paired with neutral sounds (gray). Congruent and incongruent trials did not significantly differ from each other. Error bars represent standard error. $^{*}p < 0.05$, $^{***}p < 0.001$.

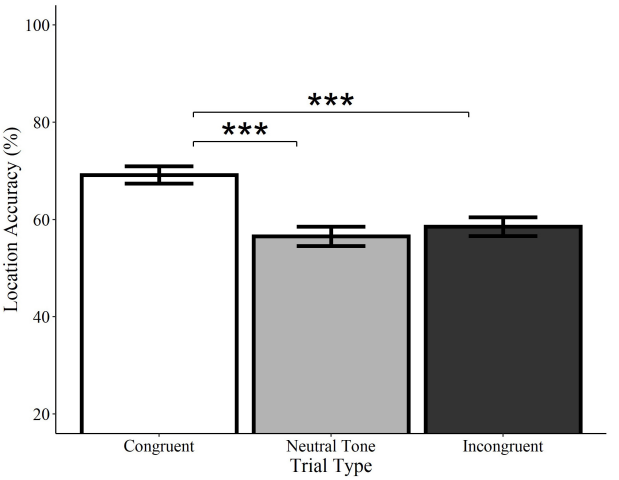

FIGURE 7 | Effect of trial type on location accuracy for environmental sounds. Locations of visual objects initially paired with congruent (white) environmental sounds were remembered with significantly greater accuracy than those paired with neutral (gray) or incongruent (black) sounds. Neutral and incongruent trials did not significantly differ from each other. Error bars represent standard error. $^{***}p < 0.001$.

Comparison of Spoken Words and Environmental Sounds

Two additional models were constructed to directly compare the impact of spoken words and environmental sounds on recognition and location accuracy[6]. In each case, accuracy was entered as the outcome variable in a generalized linear mixed-effects model with fixed effects of trial type, input, and their interaction, plus random intercepts for participant and target. Both models included word frequency, concreteness, imageability, and familiarity of the targets' labels as covariates. Trial type was treatment coded to compare congruent [0] to incongruent [1] and neutral (tone) trials [1] and input was treatment coded to compare words [0] to sounds [1]. A follow-up analysis compared incongruent trials [0] to neutral tone trials [1].

Recognition accuracy was significantly higher on congruent trials relative to neutral trials (*Estimate* $= -0.33$, $SE = 0.14$, $z = -2.40$, $p = 0.017$), as well as on incongruent trials relative to neutral trials (*Estimate* $= -0.30$, $SE = 0.14$, $z = -2.18$, $p = 0.029$). There was no main effect of input or interactions between input and any of the trial type contrasts ($ps > 0.05$).

For location accuracy, there was a significant interaction between input and the comparison of congruent to incongruent

[6]Models comparing words to sounds included all congruent and incongruent trials, as well as the neutral tone trials. Neutral word trials exclusive to the spoken word task were excluded from the analyses.

trials (*Estimate* = −0.48, *SE* = 0.18, *z* = −2.55, *p* = 0.011), as well as between input and the comparison of congruent to neutral trials (*Estimate* = −0.41, *SE* = 0.19, *z* = −2.15, *p* = 0.032; see above), confirming that congruent sounds, but not words, enhanced memory for object locations. No interaction was found between input and the comparison of incongruent trials to neutral trials (*Estimate* = 0.07, *SE* = 0.19, z = 0.37, *p* = 0.712) and there were no main effects of trial type or input (*ps* > 0.05).

DISCUSSION

The present experiment was designed to examine the extent and nature of audio-visual interactivity in visual memory, as well as the role of experience-dependent associations between cross-modal stimuli. Memory for objects' identities (i.e., "what") revealed that listening to meaningful spoken words or environmental sounds both enhanced memory for *what* visual objects were previously seen (relative to a neutral tone), regardless of whether or not the sounds were semantically congruent with the visual stimulus. One possibility is that the enhanced recognition memory in response to incongruent cues (relative to neutral cues) may be driven by expectancy violations, which have been shown to engage attentional processes (Parmentier et al., 2011; Vachon et al., 2012). Within the auditory modality, there is evidence to suggest that, under some conditions, environmental sounds may be recognized more accurately when embedded in an *incongruent* auditory context (e.g., the sound of cow mooing in a soundscape of a bowling alley) compared to a congruent auditory scene (e.g., sounds of a farm; Leech et al., 2009; Gygi and Shafiro, 2011). Gygi and Shafiro (2011) conjecture that such an Incongruency Advantage for recognition accuracy may arise from the relative novelty of incongruent stimuli and the sensitivity of sensory systems to contrasting events (Kluender et al., 2003; Ulanovsky et al., 2003). Similarly, Loftus and Mackworth (1978) found that participants studying images in preparation for a recognition memory task allocated greater visual attention to objects that were incongruent with the visual context (e.g., a picture of an octopus in a farm scene) than to objects that were contextually congruent (e.g., a picture of a tractor in a farm scene).

Alternatively (or additionally), the finding that both congruent and incongruent auditory inputs facilitated recognition memory relative to neutral tones may indicate that, regardless of the match between the visual and auditory objects, exposure to meaningful auditory stimuli initiates deeper or more elaborate semantic processing of the visual object that extends beyond its perceptual features (Craik and Lockhart, 1972; Craik and Tulving, 1975; Humphreys and Chalmers, 2016). For instance, Craik and Tulving (1975) found that recognition memory for previously seen words is enhanced following semantic judgments of the linguistic stimuli (e.g., category membership) compared to evaluations of their perceptual or phonological features (e.g., typescript and rhymes). Though participants in the present study were instructed to ignore the auditory stimuli, the enhanced memory for visual objects paired with meaningful words and sounds suggests that participants did engage in some level of auditory processing, and that meaningful auditory stimuli may promote more elaborate semantic processing of concurrently presented visual objects as well.

Importantly, however, there is reason to expect that basic semantic elaboration may have a more significant impact on memory for objects' identities, which may be encoded semantically (e.g., "I saw a cat"; Wolfe, 1998; Konkle et al., 2010) and/or perceptually (e.g., the visual memory of a cat) than on memory for objects' locations, which may rely more extensively on encoding of episodic perceptual details. Such an explanation could help account for the fact that recognition memory was facilitated by meaningful auditory inputs regardless of semantic congruency, while memory for *where* objects were previously seen was selectively enhanced by concurrent presentation of a semantically congruent, but not incongruent environmental sound or either congruent or incongruent spoken words. Marks (1989) found that semantic elaboration of pictures and their labels facilitated later recall of picture names, but not perceptual details. This finding is consistent with the view that visual objects can be encoded into memory via distinct semantic and perceptual (or episodic) pathways (e.g., dual-coding theory; Paivio, 1971, 1986), and that semantic elaboration may have a more robust impact on the former than the latter. There is additionally evidence that semantic dementia (Boxer et al., 2003; Davies et al., 2004; Moss et al., 2005; Patterson et al., 2007) and damage to areas that support semantic memory (e.g., subregions of the anterior temporal cortex; Bowles et al., 2007) disproportionately impact familiarity-based memory and object recognition relative to recollection-based episodic memory, which have been shown to be functionally and anatomically dissociable (see Brown and Aggleton, 2001; Eichenbaum et al., 2007; Ranganath and Ritchey, 2012 for reviews). For instance, Ranganath and Ritchey (2012) review evidence showing that the perirhinal cortex plays a substantial role in both familiarity-based recognition memory (e.g., for objects) and semantic processing, while the parahippocampal cortex is especially critical for the recollection of spatial and contextual details. To the extent that both congruent and incongruent auditory inputs can initiate deeper semantic processing of a concurrently presented visual object, exposure to either type of meaningful cue may subsequently facilitate recognition memory for the visual object's identity.

In contrast, we theorize that the selective facilitation of spatial memory by congruent environmental sounds may primarily stem from effects of multisensory processing on *episodic* memory, and in particular, the formation of multisensory object-based representations (Kahneman et al., 1992) and their associated contexts (i.e., "event files"; Hommel, 2004). Eye-tracking studies have shown that when participants are prompted to recall information associated with a previously seen object, they will often make visual fixations to the object's prior location, even if it is no longer visible (i.e., the "looking at nothing" phenomenon; Ferreira et al., 2008). This suggests that episodic visual memories are encoded along with spatial information, both of which may be reactivated during retrieval. Such effects have been

observed both when participants are asked to recall features of the visual object itself (e.g., Umar et al., 2021), as well as when they are asked to recall auditory stimuli presented along with a visual stimulus (e.g., Hoover and Richardson, 2008). Critically, however, object-based encoding and attention appear to be highly contingent on causal relationships and spatiotemporal continuity among different features (Pylyshyn and Storm, 1988; Hoover and Richardson, 2008). For instance, Hoover and Richardson (2008) found that when participants were asked to recall semantic facts spoken to them by an animated rabbit, they made approximately equivalent numbers of fixations to the position where the rabbit had initially communicated the auditory information, as well as to a second location where the same rabbit reappeared after being shown to burrow underground between the two mounds. Importantly, participants did not preferentially fixate the second location when an identical rabbit appeared there from a different location off-screen, demonstrating that object-based binding of visual, auditory, and spatial inputs depends on real-world constraints. It may therefore be the case that object-based memory traces of visual objects and their spatial positions may be strengthened by the concurrent presentation of semantically congruent, but not incongruent sounds, particularly when the auditory cue is typically a reliable indicator of its referent's physical location (i.e., environmental sounds, but not spoken words). Together, these findings suggest that hearing and seeing characteristics of the same object can facilitate visual memory, with the impact of auditory stimuli varying as a function of prior experiences with particular types of input.

Audio-Visual Interactions in Visual Memory

There is now considerable evidence that even basic sensory processes can be impacted by cross-modal and top-down influences through lateral and descending pathways (see Macaluso and Driver, 2005; Driver and Noesselt, 2008 for reviews). For instance, Marian et al. (2018b) demonstrated that the brain can exert top-down control over the amplification of speech sounds in the cochlea, and does so selectively depending on whether complementary visual cues are available to aid in comprehension. Particularly relevant to the present investigation, prior work has demonstrated that spatially uninformative auditory cues (e.g., a tone) can increase attentional capture and the detection of visual targets (Vroomen and De Gelder, 2000; Van der Burg et al., 2008; Matusz and Eimer, 2011). For instance, Matusz and Eimer (2011) observed that detection of a visual target (e.g., a horizontal blue bar) was facilitated when a color-change cue in the same location was accompanied by a tone relative to when the visual cue was presented unimodally.

To date, however, the majority of studies investigating audio-visual interactivity has been restricted to cross-modal interactions during perceptual and attentional tasks. The present findings indicate that cross-modal interactions during the initial processing of complex, naturalistic objects modulate how visual information is subsequently remembered.

The finding that the concurrent presentation of meaningful auditory and visual cues enhances recognition of visual objects is consistent with prior work demonstrating facilitative effects of multisensory encoding (Murray et al., 2004, 2005; Lehmann and Murray, 2005; Thelen et al., 2012, 2014, 2015; Moran et al., 2013; Thelen and Murray, 2013; Heikkilä et al., 2015; Matusz et al., 2015, 2017; Ueno et al., 2015). Previously, this type of memory enhancement has often been attributed to cross-modal interactions during the *retrieval* of visual information from memory, such that re-exposure to a previously-seen visual object initiates rapid re-activation of corresponding perceptual experiences. For instance, prior work on multisensory memory has demonstrated that the retrieval of visual and auditory information is associated with similar neural activation patterns observed during the perception of multimodal stimuli (e.g., the visual and auditory cortex, respectively; Wheeler et al., 2000). Memory for *what* objects were previously seen can therefore benefit from two sources of relevant information – the auditory and visual memory traces pointing to the same object. While such an explanation implies a high degree of interactivity between auditory and visual representations stored in memory, it does not speak to the question of whether exposure to cross-modal sensory inputs changes how memories are encoded *within* a given modality (e.g., an effect of auditory input on visual memory). The results of the present study provide support for this possibility by showing that audio-visual processing can enhance memory for information encoded exclusively by the visual system. Given that the auditory inputs in the present study did not contain relevant information regarding where an object was previously seen, their facilitation of visuospatial memory suggests that cross-modal interactivity may modulate visual memory.

The Role of Experience in Audio-Visual Interactions

In addition to demonstrating that auditory processing can enhance visual memory in the absence of redundant cues, the results of the present study suggest that the nature of cross-modal interactivity varies as a function of prior experience with particular forms of auditory and visual stimuli. Consistent with the observation that cross-modal facilitation is greater for combinations of audio-visual stimuli that commonly co-occur in naturalistic contexts (Iordanescu et al., 2011), the extent to which auditory stimuli can facilitate later memory for objects' locations may depend on how reliably the spatial location of a given sound correlates with that of its associated referent during real-world experiences.

Characteristic sounds of objects are, by their very nature, physically tied to their source, and orienting to the location of an object's sound is very likely to provide information about its visual properties. The source of spoken words, on the other hand, is often spatially displaced from that of its physical referent, making it an unreliable cue for the object's location. Our discovery that environmental sounds are more likely to facilitate visual memory for object locations than spoken words (even when neither auditory cue is spatially

informative) is consistent with the idea that the impact of auditory stimuli on visuospatial memory emerges as a result of experience-dependent changes to how the cognitive system responds to particular types of information. Furthermore, the present findings demonstrate that the processes engaged during real-world listening conditions (under which environmental sounds typically contain meaningful spatial information) persist even when auditory cues are presented monophonically and dissociated from the location of their visual referents.

One point to note, however, when interpreting the observed differences between spoken words and environmental sounds is that the two conditions differed in the proportion of congruent, incongruent, and neutral trials. Specifically, because the spoken word condition included two types of neutral trials (pseudowords and meaningless tones), congruent trials represented 1/4 of all spoken word trials, as compared to 1/3 of all environmental sound trials (which only included neutral tones). The greater advantage for spatial memory observed in response to congruent sounds compared to words could therefore be (at least in part) attributed to the fact that the congruent sounds constituted more reliable cues. The effects of trial type on *recognition* memory, however, were remarkably consistent across inputs, which speaks against this alternative explanation. To the extent that the higher proportion of neutral trials (relative to semantically congruent and incongruent trials) in the spoken word (vs. environmental sound) condition attenuated the semantic congruency effect on spatial memory, we would have expected to see a comparable reduction in the recognition memory advantage for meaningful (congruent/incongruent) vs. meaningless (neutral) cues. Instead, we observed that recognition memory was facilitated by meaningful words and sounds to a comparable degree relative to neutral cues. Furthermore, given that the ratio of congruent-to-incongruent trials was equivalent across input conditions, it is unlikely that a higher proportion of neutral trials would modulate the relative impact of congruent vs. incongruent cues. Nonetheless, while it is clear that congruent environmental sounds can facilitate memory for objects' spatial locations, the relative advantage of congruent sounds over words should be confirmed in future studies using equal proportions of congruent, incongruent, and neutral trials across the two input conditions. Future studies would additionally benefit from assessing the rate and accuracy of identification for spoken word vs. environmental sound stimuli. For instance, our finding that spatial memory was enhanced in response to congruent environmental sounds (but not congruent words or neutral/incongruent sounds and words) could conceivably be attributed to greater identifiability of sound vs. word stimuli. Given that prior research indicates that environmental sounds are typically recognized at comparable, or even lower rates relative to spoken words (Uddin et al., 2018; Bartolotti et al., 2020), however, the advantage for congruent environmental sounds observed in the present study is unlikely to be attributable to greater recognition of sounds vs. words. We note that the identifiability of items within a given block (words and sounds) is unlikely to account for effects of condition (congruent, incongruent, and neutral), as lists were counterbalanced across participants so that each item was presented in each of the conditions (i.e., as the congruent and incongruent auditory stimulus, as well as the visual target).

Our findings reveal a close link between the sights and sounds of memory, evident in the enhancement of visuospatial memory by auditory experience – a finding consistent with the well-established impact of prior experience on perceptual processing. For instance, long-term experience with cognitively and perceptually demanding activities like music (e.g., Bidelman, 2016) and bilingualism (e.g., Marian et al., 2018a; Bidelman and Heath, 2019) can impact susceptibility to perceptual illusions. These include audio-visual illusions in which visual inputs bias the perceived location (e.g., the Ventriloquist Effect; Choe et al., 1975; see Vroomen and De Gelder, 2004) or identity (e.g., the McGurk Effect; McGurk and MacDonald, 1976; see Tiippana, 2014) of auditory input, as well as those characterizing effects of auditory signals on visual perception (e.g., the Double-Flash Illusion; Shams et al., 2000; see Keil, 2020). In fact, even short-term training with audio-visual stimuli can influence unimodal processing within the auditory system (e.g., Hazan et al., 2005; Song et al., 2008; Moradi et al., 2017), as well as the visual system (e.g., Eberhardt et al., 2014; Setti et al., 2014).

CONCLUSION

In sum, our coherent perception of the world relies on the brain's ability to continuously learn and predict relationships between cross-modal stimuli – those streaming in from the external environment, as well as those stored in memory based on prior experiences. Far from a modular view of the mind (Fodor, 1983), it is now clear that information derived from different modalities is used to guide even the most basic sensory processes (Churchland, 1988; Uttal, 2001; Prinz, 2006; Marian et al., 2018b; Spence, 2020). Experience with visual and auditory stimuli can have a bi-directional impact on perception and memory, where what we hear will influence what we see, what we see will influence what we hear, and what we perceive will contribute to our memory and mental models of the world. Consistent with this iterative view of cross-modal interaction, we find that listening to meaningful sounds can enhance memory for the identity and location of visual objects, and propose that visual memory may be influenced by bottom-up processing of audio-visual input, as well as top-down effects of audio-visual experience. We conclude that cross-modal interactivity in the cognitive architecture generates a cycle in which experience shapes memory and memory shapes experience.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article are available by the corresponding author upon request.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the University Institutional Review Board.

The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

VM was responsible for project administration and funding acquisition. SS and VM conceptualized and designed the experiments. SS and research assistants collected the data. SH and SS analyzed the data. SH drafted and revised the manuscript. VM and SS provided critical feedback on the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This research was supported in part by the Eunice Kennedy Shriver National Institute of Child Health and Human Development of the National Institutes of Health under Award Number R01HD059858 to VM. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2021.661477/full#supplementary-material

Supplementary Table 1 | Mean word frequency, concreteness, familiarity and imageability of word stimulus lists. Word frequency (zipf) was calculated based on SUBTLEXUS (Brysbaert and New, 2009), and concreteness, familiarity, and imageability ratings were taken from the MRC Psycholinguistic Database (Coltheart, 1981). Numbers in parentheses represent standard deviations.

Supplementary Table 2 | Mean word frequency, concreteness, familiarity and imageability of labels in sound stimulus lists.

Supplementary Table 3 | Comparison of word frequency, concreteness, familiarity, and imageability of labels in word and sound lists.

Supplementary Table 4 | Number of items per semantic category in the spoken word and environmental sound blocks in the main and follow-up (matched) analyses. Items in the follow-up matched analyses were chosen by including all items from the block containing the fewest objects per category (e.g., "ear," "finger," "skeleton" from *anatomy* in the spoken word block) and selecting a corresponding object from the more populated block that most closely resembled each object in the less populated block (e.g., "nose," "lips," "heart," respectively in the environmental sound block).

REFERENCES

Barr, D. J., Levy, R., Scheepers, C., and Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: keep it maximal. *J. Mem. Lang.* 68, 255–278. doi: 10.1016/j.jml.2012.11.001

Bartolotti, J., and Marian, V. (2012). Language learning and control in monolinguals and bilinguals. *Cogn. Sci.* 36, 1129–1147. doi: 10.1111/j.1551-6709.2012.01243.x

Bartolotti, J., Schroeder, S. R., Hayakawa, S., Rochanavibhata, S., Chen, P., and Marian, V. (2020). Listening to speech and non-speech sounds activates phonological and semantic knowledge differently. *Q. J. Exp. Psychol.* 73, 1135–1149. doi: 10.1177/1747021820923944

Beauchamp, M. S., Argall, B. D., Bodurka, J., Duyn, J. H., and Martin, A. (2004a). Unraveling multisensory integration: patchy organization within human STS multisensory cortex. *Nat. Neurosci.* 7, 1190–1192. doi: 10.1038/nn1333

Beauchamp, M. S., Lee, K. E., Argall, B. D., and Martin, A. (2004b). Integration of auditory and visual information about objects in superior temporal sulcus. *Neuron* 41, 809–823. doi: 10.1016/s0896-6273(04)00070-4

Bidelman, G. M. (2016). Musicians have enhanced audiovisual multisensory binding: experience-dependent effects in the double-flash illusion. *Exp. Brain Res.* 234, 3037–3047. doi: 10.1007/s00221-016-4705-6

Bidelman, G. M., and Heath, S. T. (2019). Neural correlates of enhanced audiovisual processing in the bilingual brain. *Neuroscience* 401, 11–20. doi: 10.1016/j.neuroscience.2019.01.003

Bolognini, N., Frassinetti, F., Serino, A., and Làdavas, E. (2005). "Acoustical vision" of below threshold stimuli: interaction among spatially converging audiovisual inputs. *Exp. Brain Res.* 160, 273–282. doi: 10.1007/s00221-004-2005-z

Bowles, B., Crupi, C., Mirsattari, S. M., Pigott, S. E., Parrent, A. G., Pruessner, J. C., et al. (2007). Impaired familiarity with preserved recollection after anterior temporal-lobe resection that spares the hippocampus. *Proc. Natl. Acad. Sci. U.S.A.* 104, 16382–16387. doi: 10.1073/pnas.0705273104

Boxer, A. L., Rankin, K. P., Miller, B. L., Schuff, N., Weiner, M., Gorno-Tempini, M. L., et al. (2003). Cinguloparietal atrophy distinguishes Alzheimer disease from semantic dementia. *Arch. Neurol.* 60, 949–956. doi: 10.1001/archneur.60.7.949

Brown, M. W., and Aggleton, J. P. (2001). Recognition memory: what are the roles of the perirhinal cortex and hippocampus? *Nat. Rev. Neurosci.* 2, 51–61. doi: 10.1038/35049064

Brysbaert, M., and New, B. (2009). Moving beyond Kučera and Francis: a critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behav. Res. Methods* 41, 977–990. doi: 10.3758/brm.41.4.977

Calvert, G. A. (2001). Crossmodal processing in the human brain: insights from functional neuroimaging studies. *Cereb. Cortex* 11, 1110–1123. doi: 10.1093/cercor/11.12.1110

Calvert, G. A., Campbell, R., and Brammer, M. J. (2000). Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Curr. Biol.* 10, 649–657. doi: 10.1016/s0960-9822(00)00513-3

Cappe, C., Thut, G., Romei, V., and Murray, M. M. (2009). Selective integration of auditory-visual looming cues by humans. *Neuropsychologia* 47, 1045–1052. doi: 10.1016/j.neuropsychologia.2008.11.003

Chen, Y. C., and Spence, C. (2010). When hearing the bark helps to identify the dog: semantically-congruent sounds modulate the identification of masked pictures. *Cognition* 114, 389–404. doi: 10.1016/j.cognition.2009.10.012

Choe, C. S., Welch, R. B., Gilford, R. M., and Juola, J. F. (1975). The "ventriloquist effect": visual dominance or response bias? *Percept. Psychophys.* 18, 55–60. doi: 10.3758/bf03199367

Churchland, P. M. (1988). Perceptual plasticity and theoretical neutrality: a reply to Jerry Fodor. *Philos. Sci.* 55, 167–187. doi: 10.1086/289425

Coltheart, M. (1981). The MRC psycholinguistic database. *Q. J. Exp. Psychol. A* 33, 497–505. doi: 10.1080/14640748108400805

Craik, F. I., and Lockhart, R. S. (1972). Levels of processing: a framework for memory research. *J. Verb. Learn. Verb. Behav.* 11, 671–684.

Craik, F. I., and Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *J. Exp. Psychol. Gen.* 104, 268–294. doi: 10.1037/0096-3445.104.3.268

Darrow, K. N., Maison, S. F., and Liberman, M. C. (2006). Cochlear efferent feedback balances interaural sensitivity. *Nat. Neurosci.* 9, 1474–1476. doi: 10.1038/nn1807

Davies, R. R., Graham, K. S., Xuereb, J. H., Williams, G. B., and Hodges, J. R. (2004). The human perirhinal cortex and semantic memory. *Eur. J. Neurosci.* 20, 2441–2446. doi: 10.1111/j.1460-9568.2004.03710.x

De Meo, R., Murray, M. M., Clarke, S., and Matusz, P. J. (2015). Top-down control and early multisensory processes: chicken vs. egg. *Front. Integr. Neurosci.* 9:17. doi: 10.3389/fnint.2015.00017

Delano, P. H., Elgueda, D., Hamame, C. M., and Robles, L. (2007). Selective attention to visual stimuli reduces cochlear sensitivity in chinchillas. *J. Neurosci.* 27, 4146–4153. doi: 10.1523/jneurosci.3702-06.2007

Driver, J., and Noesselt, T. (2008). Multisensory interplay reveals crossmodal influences on 'sensory-specific'brain regions, neural responses, and judgments. *Neuron* 57, 11–23. doi: 10.1016/j.neuron.2007.12.013

Driver, J., and Spence, C. (1998). Attention and the crossmodal construction of space. *Trends Cogn. Sci.* 2, 254–262. doi: 10.1016/s1364-6613(98)01188-7

Dunlosky, J., Hunt, R. R., and Clark, E. (2000). Is perceptual salience needed in explanations of the isolation effect? *J. Exp. Psychol. Learn. Mem. Cogn.* 26, 649–657. doi: 10.1037/0278-7393.26.3.649

Eberhardt, S. P., Auer, E. T. Jr., and Bernstein, L. E. (2014). Multisensory training can promote or impede visual perceptual learning of speech stimuli: visual-tactile vs. visual-auditory training. *Front. Hum. Neurosci.* 8:829. doi: 10.3389/fnhum.2014.00829

Edmiston, P., and Lupyan, G. (2015). What makes words special? Words as unmotivated cues. *Cognition* 143, 93–100. doi: 10.1016/j.cognition.2015.06.008

Eichenbaum, H., Yonelinas, A. P., and Ranganath, C. (2007). The medial temporal lobe and recognition memory. *Ann. Rev. Neurosci.* 30, 123–152.

Evans, K. K., and Baddeley, A. (2018). Intention, attention and long-term memory for visual scenes: it all depends on the scenes. *Cognition* 180, 24–37. doi: 10.1016/j.cognition.2018.06.022

Ferreira, F., Apel, J., and Henderson, J. M. (2008). Taking a new look at looking at nothing. *Trends Cogn. Sci.* 12, 405–410. doi: 10.1016/j.tics.2008.07.007

Fodor, J. A. (1983). *The Modularity of Mind: An Essay on Faculty Psychology*. Cambridge, MA: MIT Press.

Frens, M. A., Van Opstal, A. J., and Van der Willigen, R. F. (1995). Spatial and temporal factors determine auditory-visual interactions in human saccadic eye movements. *Percept. Psychophys.* 57, 802–816. doi: 10.3758/bf03206796

Giard, M. H., and Peronnet, F. (1999). Auditory-visual integration during multimodal object recognition in humans: a behavioral and electrophysiological study. *J. Cogn. Neurosci.* 11, 473–490. doi: 10.1162/089892999563544

Gygi, B., and Shafiro, V. (2011). The incongruency advantage for environmental sounds presented in natural auditory scenes. *J. Exp. Psychol. Hum. Percept. Perform.* 37, 551–565. doi: 10.1037/a0020671

Hazan, V., Sennema, A., Iba, M., and Faulkner, A. (2005). Effect of audiovisual perceptual training on the perception and production of consonants by Japanese learners of English. *Speech Commun.* 47, 360–378. doi: 10.1016/j.specom.2005.04.007

Heikkilä, J., Alho, K., Hyvönen, H., and Tiippana, K. (2015). Audiovisual semantic congruency during encoding enhances memory performance. *Exp. Psychol.* 62, 123–130. doi: 10.1027/1618-3169/a000279

Hollingworth, A., and Henderson, J. M. (2002). Accurate visual memory for previously attended objects in natural scenes. *J. Exp. Psychol. Hum. Percept. Perform.* 28, 113–136. doi: 10.1037/0096-1523.28.1.113

Hommel, B. (2004). Event files: feature binding in and across perception and action. *Trends Cogn. Sci.* 8, 494–500. doi: 10.1016/j.tics.2004.08.007

Hoover, M. A., and Richardson, D. C. (2008). When facts go down the rabbit hole: contrasting features and objecthood as indexes to memory. *Cognition* 108, 533–542. doi: 10.1016/j.cognition.2008.02.011

Humphreys, M. S., and Chalmers, K. A. (2016). *Thinking About Human Memory*. Cambridge: Cambridge University Press.

Hunt, R. R., and Lamb, C. A. (2001). What causes the isolation effect? *J. Exp. Psychol. Learn. Mem. Cogn.* 27, 1359–1366. doi: 10.1037/0278-7393.27.6.1359

Iordanescu, L., Grabowecky, M., Franconeri, S., Theeuwes, J., and Suzuki, S. (2010). Characteristic sounds make you look at target objects more quickly. *Atten. Percept. Psychophys.* 72, 1736–1741. doi: 10.3758/app.72.7.1736

Iordanescu, L., Grabowecky, M., and Suzuki, S. (2011). Object-based auditory facilitation of visual search for pictures and words with frequent and rare targets. *Acta Psychol.* 137, 252–259. doi: 10.1016/j.actpsy.2010.07.017

Kahneman, D., Treisman, A., and Gibbs, B. J. (1992). The reviewing of object files: object-specific integration of information. *Cogn. Psychol.* 24, 175–219. doi: 10.1016/0010-0285(92)90007-o

Keil, J. (2020). Double flash illusions: current findings and future directions. *Front. Neurosci.* 14:298. doi: 10.3389/fnins.2020.00298

Kluender, K. R., Coady, J. A., and Kiefte, M. (2003). Sensitivity to change in perception of speech. *Speech Commun.* 41, 59–69. doi: 10.1016/s0167-6393(02)00093-6

Koelewijn, T., Bronkhorst, A., and Theeuwes, J. (2010). Attention and the multiple stages of multisensory integration: a review of audiovisual studies. *Acta Psychol.* 134, 372–384. doi: 10.1016/j.actpsy.2010.03.010

Konkle, T., Brady, T. F., Alvarez, G. A., and Oliva, A. (2010). Conceptual distinctiveness supports detailed visual long-term memory for real-world objects. *J. Exp. Psychol. Gen.* 139, 558–578. doi: 10.1037/a0019165

Leech, R., Gygi, B., Aydelott, J., and Dick, F. (2009). Informational factors in identifying environmental sounds in natural auditory scenes. *J. Acoust. Soc. Am.* 126, 3147–3155. doi: 10.1121/1.3238160

Lehmann, S., and Murray, M. M. (2005). The role of multisensory memories in unisensory object discrimination. *Cogn. Brain Res.* 24, 326–334. doi: 10.1016/j.cogbrainres.2005.02.005

Loftus, G. R., and Mackworth, N. H. (1978). Cognitive determinants of fixation location during picture viewing. *J. Exp. Psychol. Hum. Percept. Perform.* 4, 565–572. doi: 10.1037/0096-1523.4.4.565

Lupyan, G., and Spivey, M. J. (2010). Redundant spoken labels facilitate perception of multiple items. *Atten. Percept. Psychophys.* 72, 2236–2253. doi: 10.3758/bf03196698

Lupyan, G., and Thompson-Schill, S. L. (2012). The evocative power of words: activation of concepts by verbal and nonverbal means. *J. Exp. Psychol. Gen.* 141, 170–186. doi: 10.1037/a0024904

Macaluso, E., and Driver, J. (2005). Multisensory spatial interactions: a window onto functional integration in the human brain. *Trends Neurosci.* 28, 264–271. doi: 10.1016/j.tins.2005.03.008

Maison, S. F., and Liberman, M. C. (2000). Predicting vulnerability to acoustic injury with a noninvasive assay of olivocochlear reflex strength. *J. Neurosci.* 20, 4701–4707. doi: 10.1523/jneurosci.20-12-04701.2000

Marian, V., Hayakawa, S., Lam, T. Q., and Schroeder, S. R. (2018a). Language experience changes audiovisual perception. *Brain Sci.* 8:85. doi: 10.3390/brainsci8050085

Marian, V., Lam, T. Q., Hayakawa, S., and Dhar, S. (2018b). Spontaneous otoacoustic emissions reveal an efficient auditory efferent network. *J. Speech Lang. Hear. Res.* 61, 2827–2832. doi: 10.1044/2018_jslhr-h-18-0025

Marks, W. (1989). Elaborative processing of pictures in verbal domains. *Mem. Cogn.* 17, 662–672. doi: 10.3758/bf03202627

Matthews, W. J., Benjamin, C., and Osborne, C. (2007). Memory for moving and static images. *Psychonom. Bull. Rev.* 14, 989–993. doi: 10.3758/bf03194133

Matusz, P. J., and Eimer, M. (2011). Multisensory enhancement of attentional capture in visual search. *Psychonom. Bull. Rev.* 18, 904–909. doi: 10.3758/s13423-011-0131-8

Matusz, P. J., Thelen, A., Amrein, S., Geiser, E., Anken, J., and Murray, M. M. (2015). The role of auditory cortices in the retrieval of single-trial auditory–visual object memories. *Eur. J. Neurosci.* 41, 699–708. doi: 10.1111/ejn.12804

Matusz, P. J., Wallace, M. T., and Murray, M. M. (2017). A multisensory perspective on object memory. *Neuropsychologia* 105, 243–252. doi: 10.1016/j.neuropsychologia.2017.04.008

McDonald, J. J., Teder- Sälejärvi, W. A., and Hillyard, S. A. (2000). Involuntary orienting to sound improves visual perception. *Nature* 407, 906–908. doi: 10.1038/35038085

McGurk, H., and MacDonald, J. (1976). Hearing lips and seeing voices. *Nature* 264, 746–748. doi: 10.1038/264746a0

Meredith, M. A., and Stein, B. E. (1986). Spatial factors determine the activity of multisensory neurons in cat superior colliculus. *Brain Res.* 365, 350–354. doi: 10.1016/0006-8993(86)91648-3

Meyerhoff, H. S., and Huff, M. (2016). Semantic congruency but not temporal synchrony enhances long-term memory performance for audio-visual scenes. *Mem. Cogn.* 44, 390–402. doi: 10.3758/s13421-015-0575-6

Miller, L. M., and D'Esposito, M. (2005). Perceptual fusion and stimulus coincidence in the cross-modal integration of speech. *J. Neurosci.* 25, 5884–5893. doi: 10.1523/jneurosci.0896-05.2005

Molholm, S., Ritter, W., Javitt, D. C., and Foxe, J. J. (2004). Multisensory visual–auditory object recognition in humans: a high-density electrical mapping study. *Cereb. Cortex* 14, 452–465. doi: 10.1093/cercor/bhh007

Molholm, S., Ritter, W., Murray, M. M., Javitt, D. C., Schroeder, C. E., and Foxe, J. J. (2002). Multi-sensory auditory–visual interactions during early sensory processing in humans: a high-density electrical mapping study. *Cogn. Brain Res.* 14, 115–128. doi: 10.1016/s0926-6410(02)00066-6

Moradi, S., Wahlin, A., Hällgren, M., Rönnberg, J., and Lidestam, B. (2017). The efficacy of short-term gated audiovisual speech training for improving auditory sentence identification in noise in elderly hearing aid users. *Front. Psychol.* 8:368. doi: 10.3389/fpsyg.2017.00368

Moran, Z. D., Bachman, P., Pham, P., Cho, S. H., Cannon, T. D., and Shams, L. (2013). Multisensory encoding improves auditory recognition. *Multisens. Res.* 26, 581–592. doi: 10.1163/22134808-00002436

Morein-Zamir, S., Soto-Faraco, S., and Kingstone, A. (2003). Auditory capture of vision: examining temporal ventriloquism. *Cogn. Brain Res.* 17, 154–163. doi: 10.1016/s0926-6410(03)00089-2

Moss, H. E., Rodd, J. M., Stamatakis, E. A., Bright, P., and Tyler, L. K. (2005). Anteromedial temporal cortex supports fine-grained differentiation among objects. *Cereb. Cortex* 15, 616–627. doi: 10.1093/cercor/bhh163

Murray, M. M., Foxe, J. J., and Wylie, G. R. (2005). The brain uses single-trial multisensory memories to discriminate without awareness. *Neuroimage* 27, 473–478. doi: 10.1016/j.neuroimage.2005.04.016

Murray, M. M., Michel, C. M., De Peralta, R. G., Ortigue, S., Brunet, D., Andino, S. G., et al. (2004). Rapid discrimination of visual and multisensory memories revealed by electrical neuroimaging. *Neuroimage* 21, 125–135. doi: 10.1016/j.neuroimage.2003.09.035

Murray, M. M., Thelen, A., Thut, G., Romei, V., Martuzzi, R., and Matusz, P. J. (2016). The multisensory function of the human primary visual cortex. *Neuropsychologia* 83, 161–169. doi: 10.1016/j.neuropsychologia.2015.08.011

Nyberg, L., Habib, R., McIntosh, A. R., and Tulving, E. (2000). Reactivation of encoding-related brain activity during memory retrieval. *Proc. Natl. Acad. Sci. U.S.A.* 97, 11120–11124. doi: 10.1073/pnas.97.20.11120

Paivio, A. (1971). *Imagery and Verbal Processes*. New York, NY: Holt, Rinehart and Winston.

Paivio, A. (1986). *Mental Representation: A Dual Coding Approach*. New York, NY: Oxford University Press.

Parkhurst, D., Law, K., and Niebur, E. (2002). Modeling the role of salience in the allocation of overt visual attention. *Vis. Res.* 42, 107–123. doi: 10.1016/s0042-6989(01)00250-4

Parmentier, F. B., Elsley, J. V., Andrés, P., and Barceló, F. (2011). Why are auditory novels distracting? Contrasting the roles of novelty, violation of expectation and stimulus change. *Cognition* 119, 374–380. doi: 10.1016/j.cognition.2011.02.001

Patterson, K., Nestor, P. J., and Rogers, T. T. (2007). Where do you know what you know? The representation of semantic knowledge in the human brain. *Nat. Rev. Neurosci.* 8, 976–987. doi: 10.1038/nrn2277

Powers, A. R., Hevey, M. A., and Wallace, M. T. (2012). Neural correlates of multisensory perceptual learning. *J. Neurosci.* 32, 6263–6274. doi: 10.1523/jneurosci.6138-11.2012

Prinz, J. J. (2006). Is the mind really modular? *Contemp. Debates Cogn. Sci.* 14, 22–36.

Pylyshyn, Z. W., and Storm, R. W. (1988). Tracking multiple independent targets: evidence for a parallel tracking mechanism. *Spatial Vis.* 3, 179–197. doi: 10.1163/156856888x00122

Raij, T., Uutela, K., and Hari, R. (2000). Audiovisual integration of letters in the human brain. *Neuron* 28, 617–625. doi: 10.1016/s0896-6273(00)00138-0

Ranganath, C., and Ritchey, M. (2012). Two cortical systems for memory-guided behaviour. *Nat. Rev. Neurosci.* 13, 713–726. doi: 10.1038/nrn3338

Royal, D. W., Carriere, B. N., and Wallace, M. T. (2009). Spatiotemporal architecture of cortical receptive fields and its impact on multisensory interactions. *Exp. Brain Res.* 198, 127–136. doi: 10.1007/s00221-009-1772-y

Salthouse, T. A., Rogan, J. D., and Prill, K. A. (1984). Division of attention: age differences on a visually presented memory task. *Mem. Cogn.* 12, 613–620. doi: 10.3758/bf03213350

Salverda, A. P., and Altmann, G. (2011). Attentional capture of objects referred to by spoken language. *J. Exp. Psychol. Hum. Percept. Perform.* 37, 1122–1133. doi: 10.1037/a0023101

Santangelo, V., and Spence, C. (2007). Multisensory cues capture spatial attention regardless of perceptual load. *J. Exp. Psychol. Hum. Percept. Perform.* 33, 1311–1321. doi: 10.1037/0096-1523.33.6.1311

Schroeder, C. E., and Foxe, J. (2005). Multisensory contributions to low-level, 'unisensory' processing. *Curr. Opin. Neurobiol.* 15, 454–458. doi: 10.1016/j.conb.2005.06.008

Seitz, A. R., Kim, R., and Shams, L. (2006). Sound facilitates visual learning. *Curr. Biol.* 16, 1422–1427. doi: 10.1016/j.cub.2006.05.048

Setti, A., Stapleton, J., Leahy, D., Walsh, C., Kenny, R. A., and Newell, F. N. (2014). Improving the efficiency of multisensory integration in older adults: audio-visual temporal discrimination training reduces susceptibility to the sound-induced flash illusion. *Neuropsychologia* 61, 259–268. doi: 10.1016/j.neuropsychologia.2014.06.027

Shams, L., Kamitani, Y., and Shimojo, S. (2000). Illusions. What you see is what you hear. *Nature* 408:788. doi: 10.1038/35048669

Shams, L., and Seitz, A. R. (2008). Benefits of multisensory learning. *Trends Cogn. Sci.* 12, 411–417. doi: 10.1016/j.tics.2008.07.006

Smith, S. M., and Vela, E. (2001). Environmental context-dependent memory: a review and meta-analysis. *Psychonom. Bull. Rev.* 8, 203–220. doi: 10.3758/bf03196157

Song, J. H., Skoe, E., Wong, P. C., and Kraus, N. (2008). Plasticity in the adult human auditory brainstem following short-term linguistic training. *J. Cog. Neurosci.* 20, 1892–1902. doi: 10.1162/jocn.2008.20131

Spence, C. (2020). Senses of place: architectural design for the multisensory mind. *Cogn. Res. Princ. Implicat.* 5, 1–26.

Spivey, M. J., Tyler, M. J., Eberhard, K. M., and Tanenhaus, M. K. (2001). Linguistically mediated visual search. *Psychol. Sci.* 12, 282–286. doi: 10.1111/1467-9280.00352

Stein, B. E., London, N., Wilkinson, L. K., and Price, D. D. (1996). Enhancement of perceived visual intensity by auditory stimuli: a psychophysical analysis. *J. Cogn. Neurosci.* 8, 497–506. doi: 10.1162/jocn.1996.8.6.497

Szekely, A., Jacobsen, T., D'Amico, S., Devescovi, A., Andonova, E., Herron, D., et al. (2004). A new on-line resource for psycholinguistic studies. *J. Mem. Lang.* 51, 247–250. doi: 10.1016/j.jml.2004.03.002

Talsma, D., Senkowski, D., Soto-Faraco, S., and Woldorff, M. G. (2010). The multifaceted interplay between attention and multisensory integration. *Trends Cogn. Sci.* 14, 400–410. doi: 10.1016/j.tics.2010.06.008

Ten Oever, S., Romei, V., van Atteveldt, N., Soto-Faraco, S., Murray, M. M., and Matusz, P. J. (2016). The COGs (context, object, and goals) in multisensory processing. *Exp. Brain Res.* 234, 1307–1323. doi: 10.1007/s00221-016-4590-z

Thelen, A., Cappe, C., and Murray, M. M. (2012). Electrical neuroimaging of memory discrimination based on single-trial multisensory learning. *Neuroimage* 62, 1478–1488. doi: 10.1016/j.neuroimage.2012.05.027

Thelen, A., Matusz, P. J., and Murray, M. M. (2014). Multisensory context portends object memory. *Curr. Biol.* 24, R734–R735.

Thelen, A., and Murray, M. M. (2013). The efficacy of single-trial multisensory memories. *Multisens. Res.* 26, 483–502. doi: 10.1163/22134808-00002426

Thelen, A., Talsma, D., and Murray, M. M. (2015). Single-trial multisensory memories affect later auditory and visual object discrimination. *Cognition* 138, 148–160. doi: 10.1016/j.cognition.2015.02.003

Tiippana, K. (2014). What is the McGurk effect? *Front. Psychol.* 5:725. doi: 10.3389/fpsyg.2014.00725

Uddin, S., Heald, S. L., Van Hedger, S. C., Klos, S., and Nusbaum, H. C. (2018). Understanding environmental sounds in sentence context. *Cognition* 172, 134–143. doi: 10.1016/j.cognition.2017.12.009

Ueno, D., Masumoto, K., Sutani, K., and Iwaki, S. (2015). Latency of modality-specific reactivation of auditory and visual information during episodic memory retrieval. *Neuroreport* 26, 303–308. doi: 10.1097/wnr.0000000000000325

Ulanovsky, N., Las, L., and Nelken, I. (2003). Processing of low-probability sounds by cortical neurons. *Nat. Neurosci.* 6, 391–398. doi: 10.1038/nn1032

Umar, H., Mast, F. W., Cacchione, T., and Martarelli, C. S. (2021). The prioritization of visuo-spatial associations during mental imagery. *Cogn. Process.* 22, 227–237. doi: 10.1007/s10339-020-01010-5

Uttal, W. R. (2001). *The New Phrenology: The Limits of Localizing Cognitive Processes in the Brain*. Cambridge, MA: The MIT Press.

Vachon, F., Hughes, R. W., and Jones, D. M. (2012). Broken expectations: violation of expectancies, not novelty, captures auditory attention. *J. Exp. Psychol. Learn. Mem. Cogn.* 38, 164–177. doi: 10.1037/a0025054

Van der Burg, E., Olivers, C. N., Bronkhorst, A. W., and Theeuwes, J. (2008). Pip and pop: nonspatial auditory signals improve spatial visual search. *J. Exp. Psychol. Hum. Percept. Perform.* 34, 1053–1065. doi: 10.1037/0096-1523.34.5.1053

von Restorff, H. (1933). Über die Wirkung von Bereichsbildungen im Spurenfeld [On the effect of field formations in the trace field]. *Psychol. Forschung* 18, 299–342. doi: 10.1007/bf02409636

Vroomen, J., and De Gelder, B. (2000). Sound enhances visual perception: cross-modal effects of auditory organization on vision. *J. Exp. Psychol. Hum. Percept. Perform.* 26, 1583–1590. doi: 10.1037/0096-1523.26.5.1583

Vroomen, J., and De Gelder, B. (2004). Perceptual effects of cross-modal stimulation: ventriloquism and the freezing phenomenon. *Handb. Multisens. Process.* 3, 1–23.

Wallace, M. T. (2004). The development of multisensory processes. *Cogn. Process.* 5, 69–83.

Watkins, S., Shams, L., Tanaka, S., Haynes, J. D., and Rees, G. (2006). Sound alters activity in human V1 in association with illusory visual perception. *Neuroimage* 31, 1247–1256. doi: 10.1016/j.neuroimage.2006.01.016

Waxman, S. R., and Gelman, S. A. (2009). Early word-learning entails reference, not merely associations. *Trends Cogn. Sci.* 13, 258–263. doi: 10.1016/j.tics.2009.03.006

Wheeler, M. E., Petersen, S. E., and Buckner, R. L. (2000). Memory's echo: vivid remembering reactivates sensory-specific cortex. *Proc. Natl. Acad. Sci. U.S.A.* 97, 11125–11129. doi: 10.1073/pnas.97.20.11125

Wolfe, J. M. (1998). Visual memory: what do you know about what you saw? *Curr. Biol.* 8, R303–R304.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Marian, Hayakawa and Schroeder. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 28 July 2021
doi: 10.3389/fnins.2021.715954

Trigeminal Contributions to the Dorsal Cochlear Nucleus in Mouse

*Timothy S. Balmer† and Laurence O. Trussell**

Vollum Institute and Oregon Hearing Research Center, Oregon Health & Science University, Portland, OR, United States

The dorsal cochlear nucleus (DCN) is the first site of multisensory integration in the auditory pathway of mammals. The DCN circuit integrates non-auditory information, such as head and ear position, with auditory signals, and this convergence may contribute to the ability to localize sound sources or to suppress perceptions of self-generated sounds. Several extrinsic sources of these non-auditory signals have been described in various species, and among these are first- and second-order trigeminal axonal projections. Trigeminal sensory signals from the face and ears could provide the non-auditory information that the DCN requires for its role in sound source localization and cancelation of self-generated sounds, for example, head and ear position or mouth movements that could predict the production of chewing or licking sounds. There is evidence for these axonal projections in guinea pigs and rats, although the size of the pathway is smaller than might be expected for a function essential for a prey animals' survival. However, evidence for these projections in mice, an increasingly important species in auditory neuroscience, is lacking, raising questions about the universality of such proposed functions. We therefore investigated the presence of trigeminal projections to the DCN in mice, using viral and transgenic approaches. We found that the spinal trigeminal nucleus indeed projects to DCN, targeting granule cells and unipolar brush cells. However, direct axonal projections from the trigeminal ganglion itself were undetectable. Thus, secondary brainstem sources carry non-auditory signals to the DCN in mice that could provide a processed trigeminal signal to the DCN, but primary trigeminal afferents are not integrated directly by DCN.

Keywords: dorsal cochlear nucleus, trigeminal, granule cell, mouse, auditory

OPEN ACCESS

Edited by:
David Pérez-González, University of Salamanca, Spain

Reviewed by:
Jeffrey Garrett Mellott, Northeast Ohio Medical University, United States
Huizhong Whit Tao, University of Southern California, United States

***Correspondence:**
Laurence O. Trussell
trussell@ohsu.edu

†Present address:
Timothy S. Balmer, School of Life Sciences, Arizona State University, Tempe, AZ, United States

Specialty section:
This article was submitted to Auditory Cognitive Neuroscience, a section of the journal Frontiers in Neuroscience

Received: *27 May 2021*
Accepted: *07 July 2021*
Published: *28 July 2021*

Citation:
Balmer TS and Trussell LO (2021) Trigeminal Contributions to the Dorsal Cochlear Nucleus in Mouse. Front. Neurosci. 15:715954. doi: 10.3389/fnins.2021.715954

INTRODUCTION

Accurate sound localization is essential for an animal's survival and much of the auditory brainstem is specialized for this function. The dorsal cochlear nucleus (DCN), one of the first central targets of cochlear input, is thought to compute a sound source by integrating auditory spectral cues with multisensory (non-auditory) information regarding the position of the head and ears from motor, somatosensory, proprioceptive, and higher level auditory processing regions (Ryugo et al., 2003). However, the sources of multisensory information are not well understood, especially in mice, a species which has become an important model in auditory neuroscience.

The trigeminal pathway is likely to contribute to sound source localization. In principle, somatosensory signals from the head and face that could inform the auditory system of the current position of the jaw and ears – especially relevant to sounds source localization in animals with

mobile pinna – are transmitted into the brainstem via the trigeminal pathway. This pathway carries cutaneous mechanosensory information from the face and head via first-order neurons of the trigeminal ganglion to the trigeminal brainstem nuclei. From the brainstem, the second-order neurons extend their axons to the contralateral thalamus along the trigeminal lemniscus. Disruption of this pathway may underlie some forms of tinnitus, the phantom percept of high-frequency sound commonly referred to as "ringing" of the ears. Somatic tinnitus has been linked to pathological enhancement of trigeminal input to DCN. Injury to the multisensory pathways that are thought to send signals to DCN can lead to tinnitus (Folmer and Griest, 2003). Intriguingly, in 80% of tinnitus patients, head, jaw, and neck movements can modulate the perception of tinnitus (Levine et al., 2003). These movements cause altered activity in DCN (Lanting et al., 2010), possibly by enhancing somatosensory input via the trigeminal pathway. Identifying the neurons involved in carrying these trigeminal signals to the auditory system will be crucial to understanding the neural mechanisms of somatic tinnitus.

The pathway that trigeminal signals take to arrive at the DCN is unclear. Specifically, whether the trigeminal ganglion projects axons directly to neurons in the cochlear nucleus or whether a polysynaptic pathway via the brainstem is the sole route is not understood and may depend on the species in question. This polysynaptic trigeminal ganglion – trigeminal nucleus – cochlear nucleus pathway has been demonstrated conclusively in cats (Itoh et al., 1987), rats (Haenggeli et al., 2005), and guinea pigs (Zhou and Shore, 2004; Zhou et al., 2007; Zeng et al., 2011). A remarkable study in mice recently revealed that these trigeminal nucleus to DCN projections were necessary to reduce the response of DCN neurons to self-generated sounds (Singla et al., 2017). This work indicated that second-order trigeminal inputs are processed by the cochlear nucleus in mice, but whether first-order trigeminal ganglion inputs also project to DCN is unknown.

The direct pathway from the trigeminal ganglion to the cochlear nucleus is less well supported. One anatomical tracing study in guinea pigs reported that the trigeminal ganglion sends axons to the small cell cap at the dorsal edge of the ventral cochlear nucleus (VCN) (Shore et al., 2000). In rats, however, no projection to the cochlear nucleus was reported in a similar anatomical tracing study (Marfurt and Rajchert, 1991). Electrophysiological studies in guinea pigs report latencies between trigeminal stimulation and effects on DCN or VCN neurons of >5 ms (Shore et al., 2003; Shore, 2005), which may be monosynaptic or polysynaptic. To our knowledge, a direct projection from the trigeminal ganglion to the cochlear nucleus has not been reported in other species. Here we examined the direct and indirect trigeminal pathways in mice using newly developed viral tracing technologies and report that direct projections from the trigeminal ganglion to DCN were below the level of detection, whereas the indirect projection via brainstem trigeminal nuclei was present and therefore may underly the integration of head, jaw, face, and ear signals in the auditory system, and may be related to somatic tinnitus.

RESULTS

Trigeminal Nucleus Projections to DCN and Other Targets

To determine the projection pattern of trigeminal brainstem regions, we utilized an AAV1-Syn-Cre virus that is transferred from neurons at the injection site to their postsynaptic targets. After this monosynaptic anterograde transfer, the virus expresses Cre recombinase, which leads to the expression of a fluorescent protein in the Ai9 tdTomato reporter mouse (Zingg et al., 2017, 2020). AAV1-Syn-Cre was injected into the spinal trigeminal nucleus (SpV) of Ai9 mice (**Figure 1A**), and the axonal projections of the infected neurons were traced to their postsynaptic partners, as shown by clearly labeled tdTomato positive fibers and cell bodies in the facial motor nucleus, superior colliculus, and thalamus (**Figures 1B–D**). This pattern of expression confirms that SpV both projects to and makes synaptic contacts in these regions. While this result is expected, it verifies the efficacy of the transsynaptic labeling approach. While it is in principle possible for this approach to also label these circuit components in the reverse direction (with SpV as the postsynaptic target), such connectivity is not apparent in the Allen Brain Atlas database[1].

In the same experiments, fibers and cell bodies were also observed in the DCN and adjacent granule cell domain of the cochlear nucleus, although the density of cells was markedly less than was observed in other SpV targets (**Figure 2**). The labeled postsynaptic cells included unipolar brush cells (UBCs), characterized by their brush dendrite (**Figure 2Aii**) and granule cells, identified by their small size and short, spindly dendrites (**Figures 2Bii,Cii,Civ**). Labeling in fibers of the DCN molecular layer (**Figures 2Biii,iv**), where parallel fibers accumulate, further confirms that granule cells are a target of SpV fibers. Thus, the SpV was confirmed to provide input to the DCN and granule cell domain, consistent with previous studies in mice (Singla et al., 2017).

First-Order Trigeminal Nerve Projections to DCN

To test the hypothesis that the trigeminal ganglion projects directly to DCN and granule cell domains, we utilized an engineered adeno-associated virus (AAV.PHP-s.CAG.tdTomato) that, when injected intravenously, infects and fluorescently labels the peripheral nervous system (Chan et al., 2017). This approach was favored over attempting to inject a virus into the trigeminal ganglion because it could more homogenously and intensely label all three branches of the trigeminal nerve. Somata of the trigeminal ganglion were well-labeled and their axons that make up the trigeminal nerves were clearly seen entering the brain (**Figure 3A**). Although not all of the trigeminal ganglion neurons were labeled, the variety of soma sizes that were labeled indicates that a diverse sample of cell types was infected by this virus (**Figures 3B,C**). We note that while the trigeminal ganglion was strongly labeled by this virus, auditory and vestibular ganglia were

[1]https://connectivity.brain-map.org/

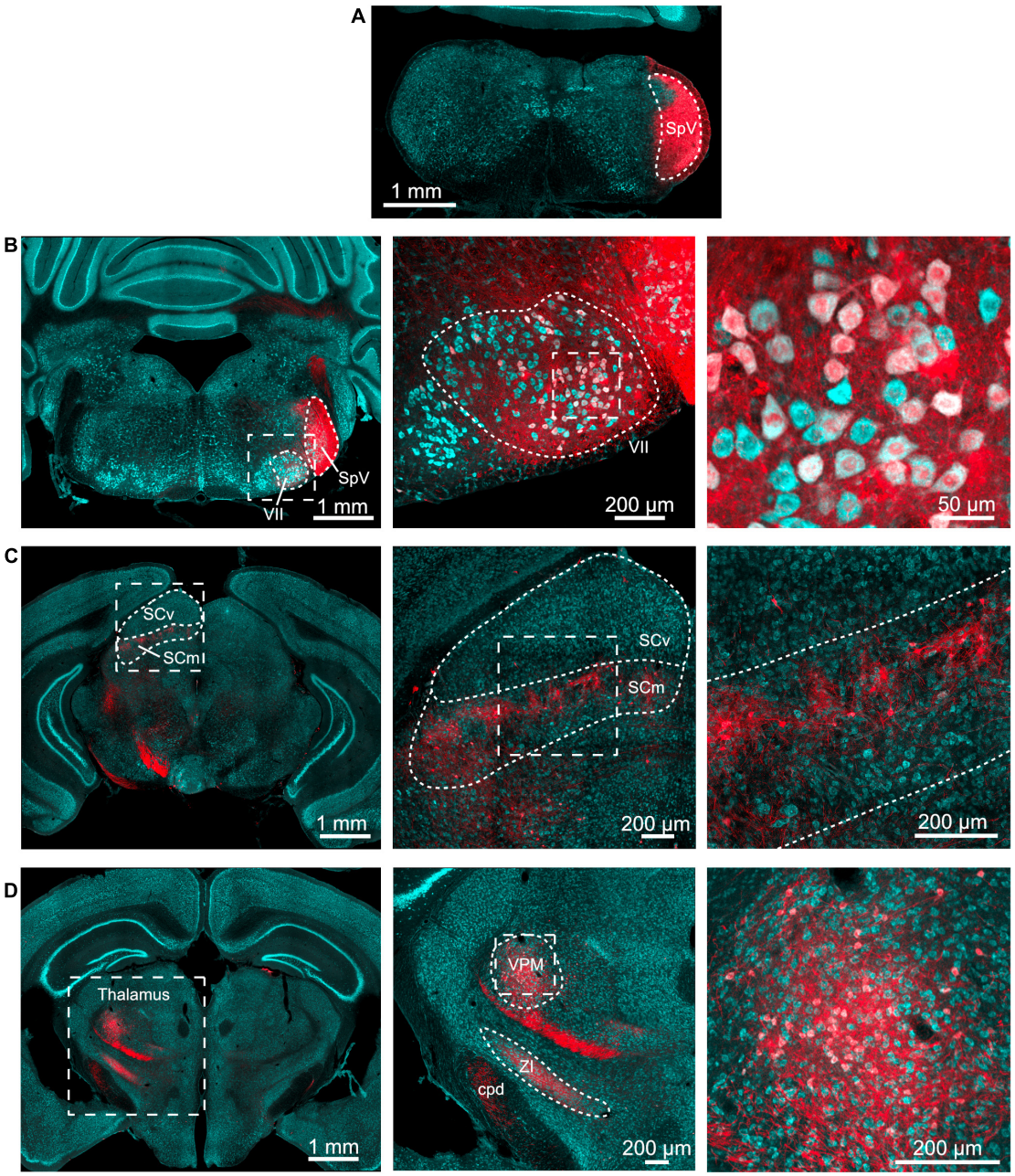

FIGURE 1 | Spinal trigeminal nucleus injection labels expected targets. **(A)** AAV1-Syn-Cre (200 nl) was injected into the SpV of tdTomato-reporter (Ai9) mice. This is the injection site. Cyan – Nissl and red – tdTomato. **(B)** Neurons in the ipsilateral facial motor nucleus (VII) were labeled. Outline of SpV shown includes the tracts to its lateral side. Middle, right, magnified view of facial motor nucleus. **(C)** Neurons in the motor-related areas of the contralateral superior colliculus were labeled. SCv, superior colliculus, visual layers; SCm, superior colliculus, multisensory layers. Middle, right, magnified views of the boxed regions in panel to left. **(D)** Neurons in the thalamus were labeled. Middle, right, magnified views of the boxed regions in panel to left. ZI, zona incerta; VPM, ventral posteromedial nucleus of the thalamus; cpd, cerebellar peduncle. All images are from the same animal, although similar results were obtained from four animals with injection volumes varying from 100 to 500 nl.

not labeled, either due to a lack of accessibility to the inner ear vasculature or due to specificity of the viral serotype.

In the brainstem, SpV had a confluence of tdTomato positive fibers throughout the brainstem, extending well into the region of the cerebellar nuclei and cerebellar cortex (**Figure 4A**). The cerebellar labeling confirmed the presence of direct trigeminal projection to the cerebellum in mice (Marfurt and Rajchert, 1991), and shows that the expression of tdTomato in trigeminal fibers reaches far into the brain. However, in marked contrast to this dense labeling, trigeminal ganglion fibers in the cochlear nucleus were undetectable (**Figures 4B,D**). We conclude that the trigeminal ganglion does not directly target the cochlear

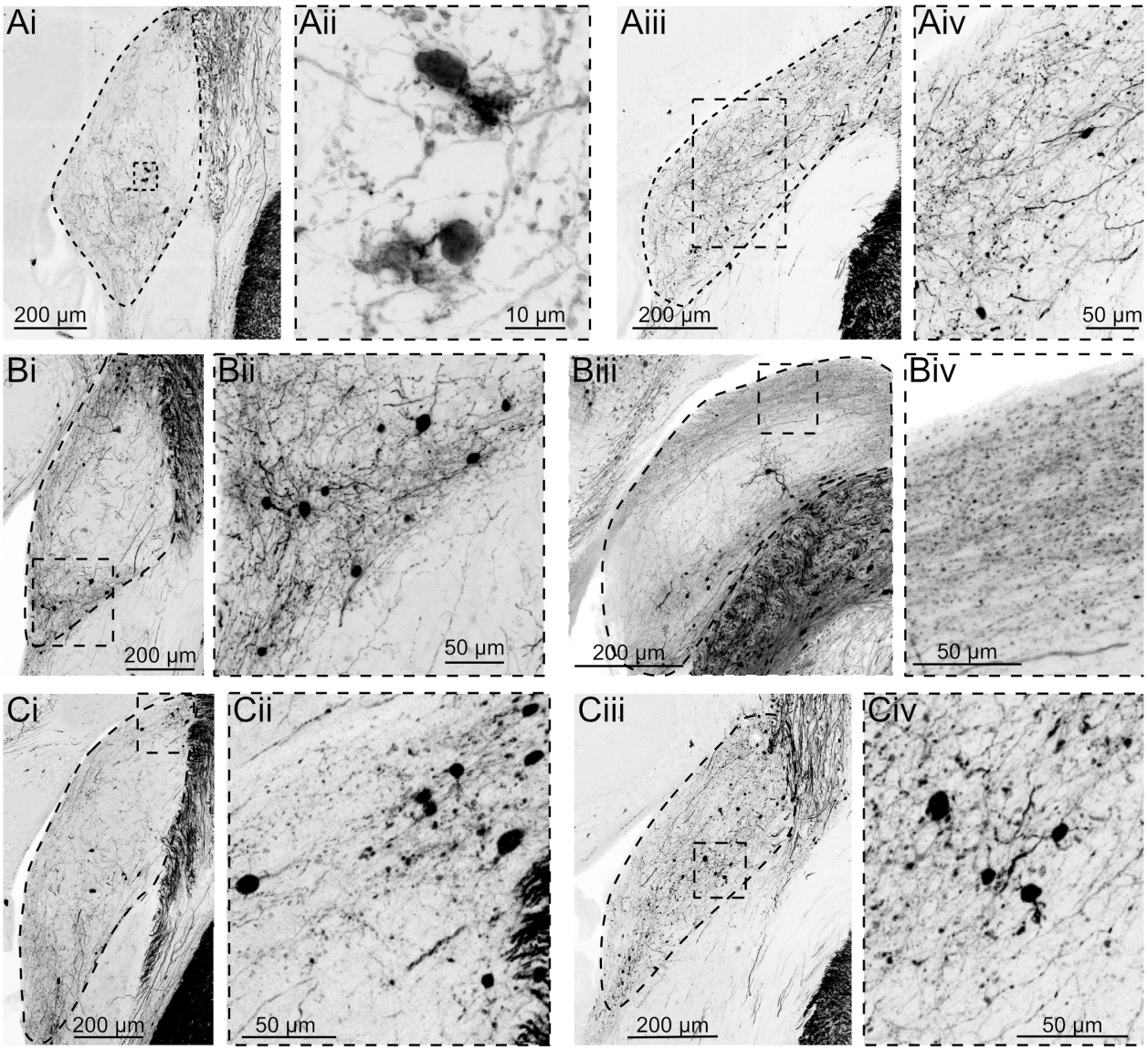

FIGURE 2 | AAV1-Syn-Cre injection into SpV reveals projections to granule domain cells in DCN. Three different experiments, each with two example coronal sections containing DCN. tdTomato labeling is shown in black. DCN is outlined in the lower magnification images. **(Ai)** Although few neurons are transsynaptically labeled, UBCs can be identified by their distinct morphology. **(Aii)** Magnification of the boxed region in **Ai**. **(Aiii)** A caudal section containing DCN from the same animal. **(Aiv)** Magnification of the boxed region in **Aiii**. **(Bi)** Another experiment showing transsynaptically labeled small cells in the region between DCN and VCN. **(Bii)** Magnification of the boxed region in **Bi**. **(Biii)** A more caudal section of DCN demonstrates labeled parallel fiber axons, confirming that granule cells were labeled transsynaptically. **(Biv)** Magnification of the boxed region in **Biii**. **(Ci–iv)** A third experiment further demonstrating that a small but consistent population of neurons receives synaptic input from SpV projections.

nucleus of mice, including VCN, DCN, or granule cell regions. Occasional labeled cell bodies, probably cartwheel cells, were observed in the DCN near the molecular layer (**Figure 4**). However, this labeling is consistent with the sparse glia and neurons labeled throughout the brain primarily near the surface of the brain or near blood vessels, presumably due to the intravenous delivery of the virus (cerebellar Purkinje cells in **Figure 4A**). This AAV.PHP-s virus has not been reported to jump across synapses and we do not suspect that these labeled cells are postsynaptic to primary afferents. The fact that the sparse labeling of cells is not due to transsynaptic spread of virus is supported by the absence of somatic labeling within the regions of the trigeminal nuclei despite the presence of dense afferent fibers (**Figure 4**).

DISCUSSION

Mouse models for hearing disorders are increasingly common, due to the ready availability of genetic mutations that affect

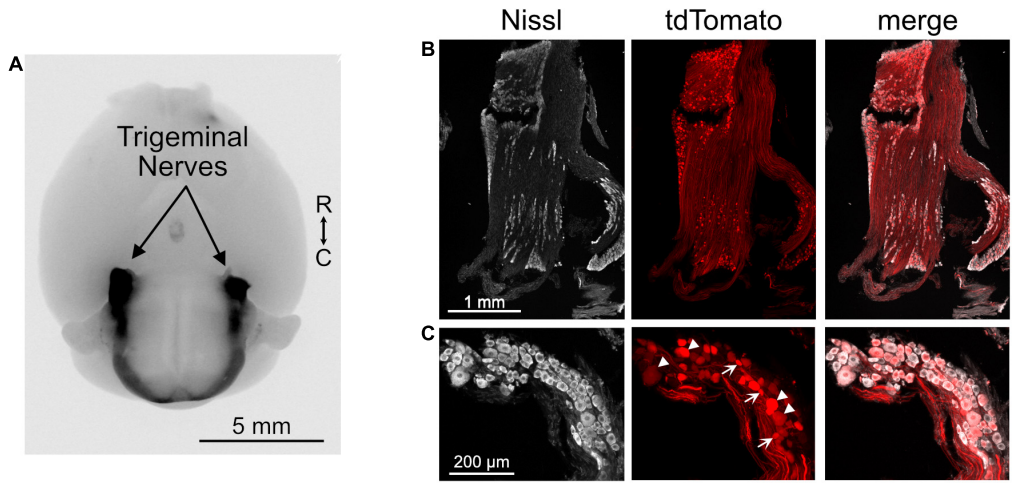

FIGURE 3 | Trigeminal ganglion labeled with AAV.PHP-s virus. **(A)** Image of ventral aspect of the brain of a mouse that received a retro-orbital injection of AAV.PHP-s.CAG.tdTomato. The trigeminal nerves are brightly labeled, as are axons in the spinal cord that likely originate from dorsal root ganglia. tdTomato is shown in black. R, rostral; C, caudal. **(B)** Section of trigeminal ganglion demonstrating that many somata were labeled by this approach. **(C)** Neurons with a diverse range of soma sizes were labeled. Arrowheads – large somata and arrows – small somata.

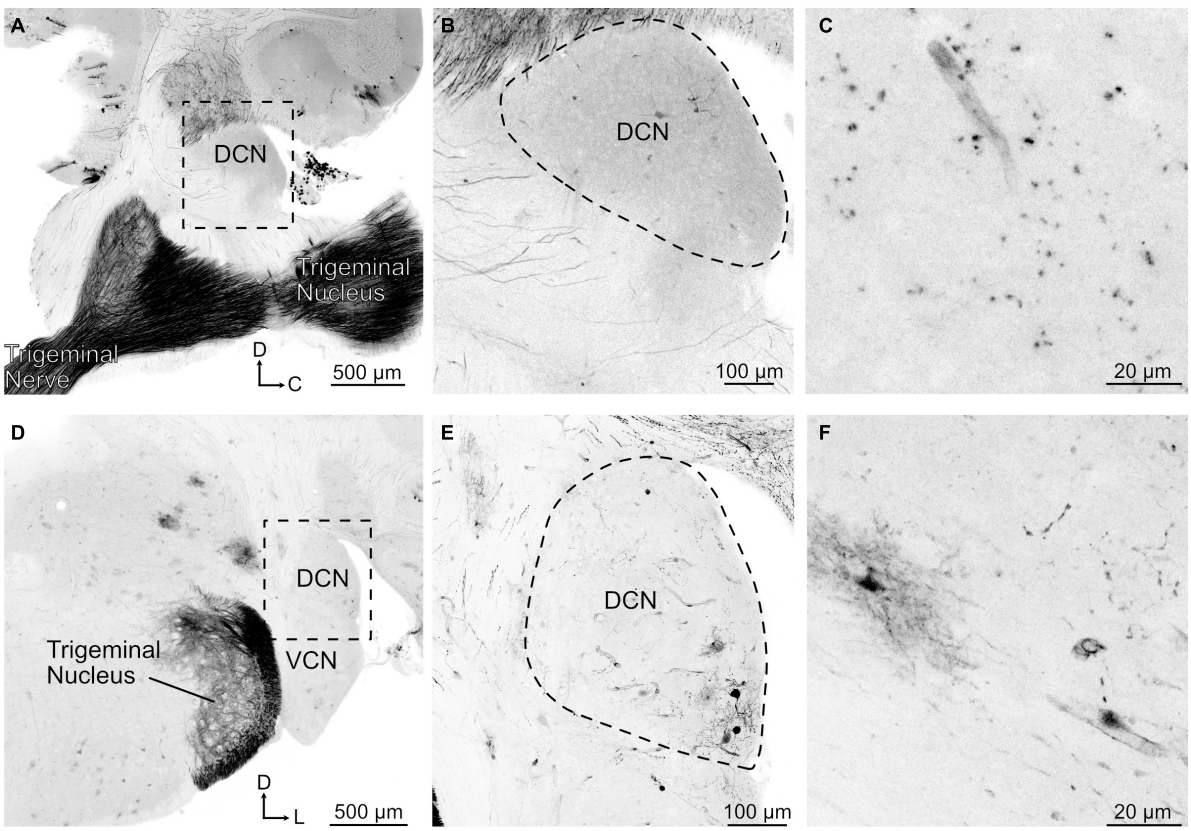

FIGURE 4 | Trigeminal nerve makes few projections to DCN. **(A)** Sagittal section containing trigeminal nerve, trigeminal nucleus, and DCN. Note that there are few if any labeled axonal fibers that project to DCN despite strong labeling of nerve. tdTomato labeling is shown in black. D, dorsal; C, caudal. **(B)** Magnified view of the boxed region in **A**. **(C)** Higher magnification of a sagittal DCN section showing labeled blood vessel. **(D)** Coronal section showing trigeminal nerve fibers labeled in the trigeminal nucleus but not in the DCN. VCN, ventral cochlear nucleus; D, dorsal; L, lateral. **(E)** Magnified view of the boxed region in **D** showing that the labeled neurons are clustered around a blood vessel. **(F)** Higher magnification of a coronal DCN section showing a labeled blood vessel, glial cell, and few neuronal processes.

hearing, or enhance the accessibility of cellular elements for anatomical or physiological study (Ohlemiller, 2019). Recent studies of tinnitus have focused on alterations in the function of principal cells of the DCN that might be related to plasticity in the inputs from non-auditory sources associated sensation in the head and neck (Levine et al., 2003; Kaltenbach, 2006). The role of trigeminal input to granule cell domains in and around the cochlear nucleus has been proposed to be part of this modulatory pathway affected during tinnitus (Shore et al., 2007). In other species, stimulation of trigeminal inputs leads to spike modulation in DCN and VCN (Shore et al., 2003; Shore, 2005), but direct evidence for this in mice is lacking. In the present work, we sought to provide further support for the use of the mouse model by confirming anatomically trigeminal projections to the cochlear nucleus. Unlike previous anatomical studies, we used a viral approach to take advantage of the enhanced infection afforded by AAV reagents in the hope that a robust labeling would be obtained.

Indeed, injection into trigeminal nuclei of tdTomato reporter mice of an anterogradely transported Cre-expressing virus capable of transsynaptic labeling led to robust expression of labeled fibers and cells in numerous known trigeminal targets, including trigeminal nuclei, thalamus, superior colliculus, facial nucleus, and cerebellum. Fibers and cell bodies, including presumptive granule cells and unipolar brush cells, were also found in the granule cell lamina and cell body region of the DCN, although at an apparently lower density than in other trigeminal targets. Such labeling is consistent with a previous study in mice showing trigeminal nuclear input to DCN and its apparent necessity for modulation of the DCN's response to self-generated sounds associated with licking (Singla et al., 2017).

In contrast to this result obtained by virus injection into a second-order nucleus, we were unable to confirm that primary trigeminal afferents project to the cochlear nucleus. Here, we used a systemically injected virus previously shown to be taken up in peripheral ganglia (Chan et al., 2017). Several observations confirmed that labeling with this approach was sufficient to label nearly all ganglionic inputs. Expression of the tdTomato reporter was obvious throughout the ganglion, the seventh cranial nerve, and its projections into the brain. Indeed, dense fibers were seen in trigeminal nuclei, as well as fibers extending into the cerebellum. The latter projection has been described previously and was considered "scant" as compared to secondary trigeminal projections (Marfurt and Rajchert, 1991), but was quite obvious in our micrographs.

However, no primary trigeminal afferent fibers were apparent in cochlear nucleus, in contrast to a previous report in guinea pig (Shore et al., 2000). We consider three reasons for these conflicting results. First, it may be that the difference in species is a factor. The hearing range of guinea pigs is somewhat lower than that of mice (Warfield, 1973; Fay, 1988), and perhaps the functional significance of convergence to the DCN from non-auditory sources varies as well. Second, it may be that the primary trigeminal input is sparse even in guinea pig, consistent with the micrographs shown in Shore et al. (2000). While trigeminal stimulation drives activity in guinea pig DCN, this response could largely reflect a disynaptic circuit. A third possibility is that the viral method we used here was somehow biased against infection of those ganglion somata that target the cochlear nucleus. However, regardless of the reasons, our results provide confirmation of trigeminal nuclear innervation of the cochlear nucleus and further support the use of the mouse as a model for somatosensory modulation of auditory function. Moreover, given that some targets of primary trigeminal afferents have also been shown to project to cochlear nucleus [dorsal horn of C1 and C2, ventral horn of cervical spinal cord, cuneate, and vestibular nuclei (Marfurt and Rajchert, 1991)], the trigeminal influence of the function of the cochlear nucleus may extend beyond the trigeminal nuclei themselves.

MATERIALS AND METHODS

Animals

Ai9(RCL-tdT) (Madisen et al., 2012) or C57BL/6J mice of both sexes were bred in-house and all procedures were approved by the Oregon Health and Science University's Institutional Animal Care and Use Committee.

Intracranial Viral Injections

Viral injections were made into the SpV of four adult mice (>3 months old) using a stereotax (Kopf), single-axis manipulator (Narishige), and pipette vice (Ronal) under isoflurane anesthesia. Glass capillaries (Drummond Scientific) were pulled on a pipette puller (Sutter P-97) and beveled at 45° angle with a 20–30 μm inside diameter using a diamond lapping disk (3M). An incision was made in the scalp along the midline, and a small hole was drilled into the skull. The pipette was lowered into the brain at 10 μm/s. Five-minute periods were allowed before and after injection. AAV1-Syn-Cre (3.15e13 GC/ml) was purchased from the University of Pennsylvania's viral vector core. In total, 100–500 nl of undiluted virus was injected using stereotaxic coordinates (7.8 mm caudal, 2.2 mm lateral, 3.5 mm ventral, relative to bregma).

Intra-Orbital Viral Injections

Two P18–30 mice were anesthetized with isoflurane and 5–30 μl of the AAV.PHP-s.CAG.tdTomato virus was injected into the retro-orbital sinus using a 0.5-ml syringe with a 28 ga needle. The AAV.PHP-s.CAG.tdTomato (1.7e13 GC/ml) was purchased form Addgene.

Immunohistochemistry and Imaging

Three weeks after viral injection, mice were overdosed with isoflurane and perfused through the heart with 0.01 M phosphate

buffered saline, 7.4 pH (PBS) followed by 4% paraformaldehyde in PBS. Brains and trigeminal ganglia were extracted from the skull and incubated in the same solution overnight at 4°C. Fifty-micrometer-thick sections were made on a vibratome and saved as floating sections in PBS. Sections were rinsed 3 × 10 min in PBS, blocked, and permeabilized in 5% normal donkey serum (NDS), 2% fish gelatin, and 0.2% Triton X-100 in PBS for >2 h at room temperature. Sections were incubated in primary antibodies to amplify the tdTomato labeling using 1:400 rabbit anti-DsRed (632496, Clontech) in 5% NDS for 2–3 days at 4°C on an orbital shaker. Sections were rinsed 3 × 10 min in PBS, followed by 1:500 donkey anti-rabbit Cy3 (711-165-153, Jackson ImmunoResearch) in 5% NDS for 2–3 days at 4°C on an orbital shaker. The sections were mounted on microscope slides and, in some cases, a fluorescent nissl stain (1:50 NeuroTrace 435/455, Invitrogen) was applied for 0.5–2 h. The slides were coverslipped with Fluoromount-G (Southern Biotech). Images were acquired on a confocal microscope (Zeiss 780 or 880) or on a Zeiss Elyra PS.1 with AiryScan system that reconstructs super-resolution images from a series of images acquired under spatially structured illumination (Gustafsson, 2000). In some cases, tdTomato labeling was converted to grayscale and then inverted in order to enhance the contrast of the fluorescent labeling. All images are maximum intensity projections.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The animal study was reviewed and approved by the Institutional Animal Care and Use Committee of OHSU.

AUTHOR CONTRIBUTIONS

LT and TB conceived the experiments and wrote the manuscript. TB conducted the experiments and analyzed the data. Both authors contributed to the article and approved the submitted version.

FUNDING

This study was funded by the National Institutes of Health (NIH) Grants NS028901 and DC004450 (PI: LT); K99 DC016905 (PI: TB); P30 DC005983 (PI: Peter Barr-Gillespie); and P30 NS0618000 (PI: Sue Aicher).

REFERENCES

Chan, K. Y., Jang, M. J., Yoo, B. B., Greenbaum, A., Ravi, N., Wu, W. L., et al. (2017). Engineered AAVs for efficient noninvasive gene delivery to the central and peripheral nervous systems. *Nat. Neurosci.* 20, 1172–1179. doi: 10.1038/nn.4593

Fay, R. R. (1988). *Hearing in Vertebrates: a Psychophysics Databook*. Winnetka: Hill-Fay Associates.

Folmer, R. L., and Griest, S. E. (2003). Chronic tinnitus resulting from head or neck injuries. *Laryngoscope* 113, 821–827. doi: 10.1097/00005537-200305000-00010

Gustafsson, M. G. (2000). Surpassing the lateral resolution limit by a factor of two using structured illumination microscopy. *J. Microsc.* 198, 82–87. doi: 10.1046/j.1365-2818.2000.00710.x

Haenggeli, C. A., Pongstaporn, T., Doucet, J. R., and Ryugo, D. K. (2005). Projections from the spinal trigeminal nucleus to the cochlear nucleus in the rat. *J. Compar. Neurol.* 484, 191–205. doi: 10.1002/cne.20466

Itoh, K., Kamiya, H., Mitani, A., Yasui, Y., Takada, M., and Mizuno, N. (1987). Direct projections from the dorsal column nuclei and the spinal trigeminal nuclei to the cochlear nuclei in the cat. *Brain Res.* 400, 145–150. doi: 10.1016/0006-8993(87)90662-7

Kaltenbach, J. A. (2006). Summary of evidence pointing to a role of the dorsal cochlear nucleus in the etiology of tinnitus. *Acta Otolaryngol. Suppl.* 2006, 20–26. doi: 10.1080/03655230600895309

Lanting, C. P., de Kleine, E., Eppinga, R. N., and van Dijk, P. (2010). Neural correlates of human somatosensory integration in tinnitus. *Hear. Res.* 267, 78–88. doi: 10.1016/j.heares.2010.04.006

Levine, R. A., Abel, M., and Cheng, H. (2003). CNS somatosensory-auditory interactions elicit or modulate tinnitus. *Exp. Brain Res.* 153, 643–648. doi: 10.1007/s00221-003-1747-3

Marfurt, C. F., and Rajchert, D. M. (1991). Trigeminal primary afferent projections to "non-trigeminal" areas of the rat central nervous system. *J. Compar. Neurol.* 303, 489–511. doi: 10.1002/cne.903030313

Madisen, L., Mao, T. Y., Koch, H., Zhuo, J. M., Berenyi, A., Fujisawa, S., et al. (2012). A toolbox of Cre-dependent optogenetic transgenic mice for light-induced activation and silencing. *Nat. Neurosci.* 15, 793–802. doi: 10.1038/nn.3078

Ohlemiller, K. K. (2019). Mouse methods and models for studies in hearing. *J. Acoust. Soc. Am.* 146, 3668–3680. doi: 10.1121/1.5132550

Ryugo, D. K., Haenggeli, C. A., and Doucet, J. R. (2003). Multimodal inputs to the granule cell domain of the cochlear nucleus. *Exp. Brain Res.* 153, 477–485. doi: 10.1007/s00221-003-1605-3

Shore, S., Zhou, J., and Koehler, S. (2007). Neural mechanisms underlying somatic tinnitus. *Prog. Brain Res.* 166, 107–23. doi: 10.1016/s0079-6123(07)66010-5

Shore, S. E. (2005). Multisensory integration in the dorsal cochlear nucleus: unit responses to acoustic and trigeminal ganglion stimulation. *Eur. J. Neurosci.* 21, 3334–3348. doi: 10.1111/j.1460-9568.2005.04142.x

Shore, S. E., El Kashlan, H., and Lu, J. (2003). Effects of trigeminal ganglion stimulation on unit activity of ventral cochlear nucleus neurons. *Neuroscience* 119, 1085–1101. doi: 10.1016/S0306-4522(03)00207-0

Shore, S. E., Vass, Z., Wys, N. L., and Altschuler, R. A. (2000). Trigeminal ganglion innervates the auditory brainstem. *J. Compar. Neurol.* 419, 271–285. doi: 10.1002/(SICI)1096-9861(20000410)419:3<271::AID-CNE1<3.0.CO;2-M

Singla, S., Dempsey, C., Warren, R., Enikolopov, A. G., and Sawtell, N. B. (2017). A cerebellum-like circuit in the auditory system cancels responses to self-generated sounds. *Nat. Neurosci.* 20, 943–950. doi: 10.1038/nn.4567

Warfield, D. (1973). "The study of hearing in animals" in *Methods of Animal Experimentation, IV*. ed. W. Gay. (United States: Academic Press). 43–143. doi: 10.1016/b978-0-12-278004-2.50008-6

Zeng, C., Shroff, H., and Shore, S. E. (2011). Cuneate and spinal trigeminal nucleus projections to the cochlear nucleus are differentially associated with vesicular glutamate transporter-2. *Neuroscience* 176, 142–151. doi: 10.1016/j.neuroscience.2010.12.010

Zhou, J., Nannapaneni, N., and Shore, S. (2007). Vessicular glutamate transporters 1 and 2 are differentially associated with auditory nerve and spinal trigeminal inputs to the cochlear nucleus. *J. Compar. Neurol.* 500, 777–787. doi: 10.1002/cne.21208

Zhou, J., and Shore, S. (2004). Projections from the trigeminal nuclear complex to the cochlear nuclei: a retrograde and anterograde tracing study in the guinea pig. *J. Neurosci. Res.* 78, 901–907. doi: 10.1002/jnr.20343

Zingg, B., Chou, X. L., Zhang, Z. G., Mesik, L., Liang, F., Tao, H. W., et al. (2017). AAV-Mediated Anterograde Transsynaptic Tagging: mapping Corticocollicular Input-Defined Neural Pathways for Defense Behaviors. *Neuron* 93, 33–47. doi: 10.1016/j.neuron.2016.11.045

Zingg, B., Peng, B., Huang, J., Tao, H. W., and Zhang, L. I. (2020). Synaptic Specificity and Application of Anterograde Transsynaptic AAV for Probing Neural Circuitry. *J. Neurosci.* 40, 3250–3267. doi: 10.1523/jneurosci.2158-19.2020

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Balmer and Trussell. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 29 July 2021
doi: 10.3389/fnins.2021.642047

Effects of Multisession Anodal Electrical Stimulation of the Auditory Cortex on Temporary Noise-Induced Hearing Loss in the Rat

*Iván Díaz, Ana Cecilia Colmenárez-Raga, David Pérez-González, Venezia G. Carmona, Ignacio Plaza Lopez and Miguel A. Merchán**

Instituto de Neurociencias de Castilla y León (INCYL), Universidad de Salamanca, Salamanca, Spain

OPEN ACCESS

Edited by:
Etienne De Villers-Sidani, McGill University, Canada

Reviewed by:
Maria Vittoria Podda, Catholic University of the Sacred Heart, Italy
Andrej Kral, Hannover Medical School, Germany

****Correspondence:***
Miguel A. Merchan
merchan@usal.es

Specialty section:
This article was submitted to Auditory Cognitive Neuroscience, a section of the journal Frontiers in Neuroscience

Received: *15 December 2020*
Accepted: *30 April 2021*
Published: *29 July 2021*

Citation:
Díaz I, Colmenarez-Raga AC, Pérez-González D, Carmona VG, Plaza Lopez I and Merchán MA (2021) Effects of Multisession Anodal Electrical Stimulation of the Auditory Cortex on Temporary Noise-Induced Hearing Loss in the Rat. Front. Neurosci. 15:642047. doi: 10.3389/fnins.2021.642047

The protective effect of the efferent system against acoustic trauma (AT) has been shown by several experimental approaches, including damage to one ear, sectioning of the olivocochlear bundle (OCB) in the floor of the IV ventricle, and knock-in mice overexpressing outer hair cell (OHC) cholinergic receptors, among others. Such effects have been related to changes in the regulation of the cholinergic efferent system and in cochlear amplification, which ultimately reverse upon protective hearing suppression. In addition to well-known circuits of the brainstem, the descending corticofugal pathway also regulates efferent neurons of the olivary complex. In this study, we applied our recently developed experimental paradigm of multiple sessions of electrical stimulation (ES) to activate the efferent system in combination with noise overstimulation. ABR thresholds increased 1 and 2 days after AT (8–16 kHz bandpass noise at 107 dB for 90 min) recovering at AT + 14 days. However, after multiple sessions of epidural anodal stimulation, no changes in thresholds were observed following AT. Although an inflammatory response was also observed 1 day after AT in both groups, the counts of reactive macrophages in both experimental conditions suggest decreased inflammation in the epidural stimulation group. Quantitative immunocytochemistry for choline acetyltransferase (ChAT) showed a significant decrease in the size and optical density of the efferent terminals 1 day after AT and a rebound at 14 days, suggesting depletion of the terminals followed by a long-term compensatory response. Such a synthesis recovery was significantly higher upon cortical stimulation. No significant correlation was found between ChAT optical density and size of the buttons in sham controls (SC) and ES/AT + 1day animals; however, significant negative correlations were shown in all other experimental conditions. Therefore, our comparative analysis suggests that cochleotopic cholinergic neurotransmission is also better preserved after multisession epidural stimulation.

Keywords: corti organ, auditory brainstem responses, quantitative immunocytochemistry, choline acetyl transferase, epidural anodal direct current stimulation, cochlear inflammatory response, cochleotopy

INTRODUCTION

The medial olivocochlear (MOC) efferent system enhances hearing sound detection throughout cochlear amplifier regulation (Guinan, 2006, 2010; Lopez-Poveda, 2018), in addition to inducing hearing suppression, as shown in the seminal study by Robert Galambos (Galambos reflex) (Galambos, 1956). More recently, compound action potential (CAP) amplitude suppression and cochlear microphonic (CM) amplitude increments have been shown when applying electrical stimulation on the floor of the IV ventricle (Elgueda et al., 2011). In line with its role in regulating hearing sensitivity, efferent system activation induces a protective effect against noise overstimulation (Handrock and Zeisberg, 1982; Patuzzi and Thompson, 1991; Zheng et al., 1997; Tong et al., 2013; Dinh et al., 2015; Boero et al., 2018). Furthermore, after showing an increased resistance against hearing loss in knock-in mice (KI; Chrna9L9′TKI, carrying a positive alpha 9-receptor point mutation), it has been suggested that MOC cholinergic neurotransmission is directly involved in minimizing noise trauma (Boero et al., 2018).

Electrophysiological evidence also shows that, despite the mechanism of self-regulation of the Galambos' reflex in the low auditory pathway, the brain cortex also controls efferent olivocochlear (OC) responses (Xiao and Suga, 2002; Terreros and Delano, 2015). Accordingly, descending corticofugal regulation of the strength of the OC reflex has been demonstrated after pharmacological blocking, cooling, or macrostimulation of the auditory cortex (AC) in animal models (León et al., 2012; Dragicevic et al., 2015; Terreros and Delano, 2015). The same effect is detected in humans after cortical epidural electrical stimulation (Fenoy et al., 2006; Perrot et al., 2006).

Short periods of noise overexposure produce reversible changes in hearing loss, known as temporary threshold shifts (TTS). In recent years, the full reversibility of TTS has been questioned after showing that long-term damage of synaptic buttons and afferent fibers persist in overstimulated animals which recover their hearing threshold (Kujawa and Liberman, 2009). These masked alterations, currently known as hidden hearing loss (HHL), can evolve into auditory alterations such as hyperacusis, tinnitus, or difficulties in sound discrimination (Liberman et al., 2016). Thus, short acoustic overexposure with reversible threshold shifts stands out as an overlooked silent alteration, increasingly prevalent in our noisy world, which lacks treatment or prevention (Delano et al., 2020).

In our previous studies, we have recently communicated that chronic anodal epidural stimulation in rats promotes AC activation with transient hearing threshold elevation, as demonstrated by auditory brainstem recordings (ABRs) (Colmenárez-Raga et al., 2019). Based on these results, we hypothesized that a multisession stimulation protocol of the AC may induce a sustained and reversible decrease in hearing sensitivity. Such an effect, also explored in this study, may be used as a potential protective intervention in hearing disorders, such as acoustic trauma or hyperacusis.

Here, we assess the effects on the inner ear of chronic epidural stimulation of the AC in an animal model of transient sound overexposure. More specifically, we analyze the effects of sound overactivation in combination with multisession AC epidural activation in the inner ear of the rat. For this purpose, we applied our previously tested protocol of AC anodal epidural activation (Colmenárez-Raga et al., 2019), followed by a sound stimulation protocol designed for TTS induction [through a single session of 107 dB at a restricted frequency band (8–16 kHz)]. In our experimental approach, the protocol for acoustic stimulation was quite similar to those previously applied by other authors for TTS induction (Kujawa and Liberman, 2009, in mice or Lobarinas et al., 2017, in rats). Ultimately, this study aims to correlate, in a TTS model, the effects on hearing sensitivity of multisession epidural anodal stimulation on the AC (measured by ABR recordings) with MOC efferent cholinergic neurotransmission in the cochlea, analyzed by quantitative immunocytochemistry of choline acetyltransferase (ChAT) in surface preparations of the organ of Corti.

MATERIALS AND METHODS

This study was conducted in strict accordance with Spanish regulations (Royal Decree 53/2013—Law 32/2007) and European Union guidelines (Directive 2010/63/EU) on the care and use of animals in biomedical research. All surgeries were performed under monitored anesthesia (respiratory rate, body temperature, and oxygen saturation), and all efforts were made to minimize suffering. In total, 28 young male Wistar rats weighing from 250 to 300 g, with normal ABR hearing thresholds, were separated into four groups and treated using the following protocols: electrode implantation without any stimulation (electrical or acoustical) (Sham controls, SC), electrically stimulated (ES), acoustic trauma (AT), and electrically stimulated followed by AT (ES/AT) (**Figure 1**). Furthermore, we assessed the short-term effects of these protocols 1 day after acoustic stimulation (day 13th of the protocol) and the corresponding long-term effects 14 days after AT (day 26th). As shown in **Figure 1**, the animal groups were organized as follows:

SC—Sham controls ($n = 12$).

ES—Electrical stimulation. Euthanized at day 14th of the protocol ($n = 3$). One of the cases of this group was dropped from further analysis due to damage detected in the deep layers of the AC.

AT—Acoustic trauma. Euthanized at short term (AT + 1 day, day 13th of the protocol) ($n = 3$) and long term (AT + 14 days, day 26th of the protocol) ($n = 3$).

ES/AT—Electrical stimulation and acoustic trauma. Euthanized at short term (ES/AT + 1 day, day 13th of the protocol) ($n = 3$), long term (ES/AT + 14 days, day 26th of the protocol) ($n = 3$).

Sham control rats were histologically processed simultaneously with the treated animals (paired processing of brain sections and inner ear surface preparations).

Surgery

Under gas anesthesia (2.5% isoflurane), rats were placed in a stereotaxic frame, surgically exposing the left temporal

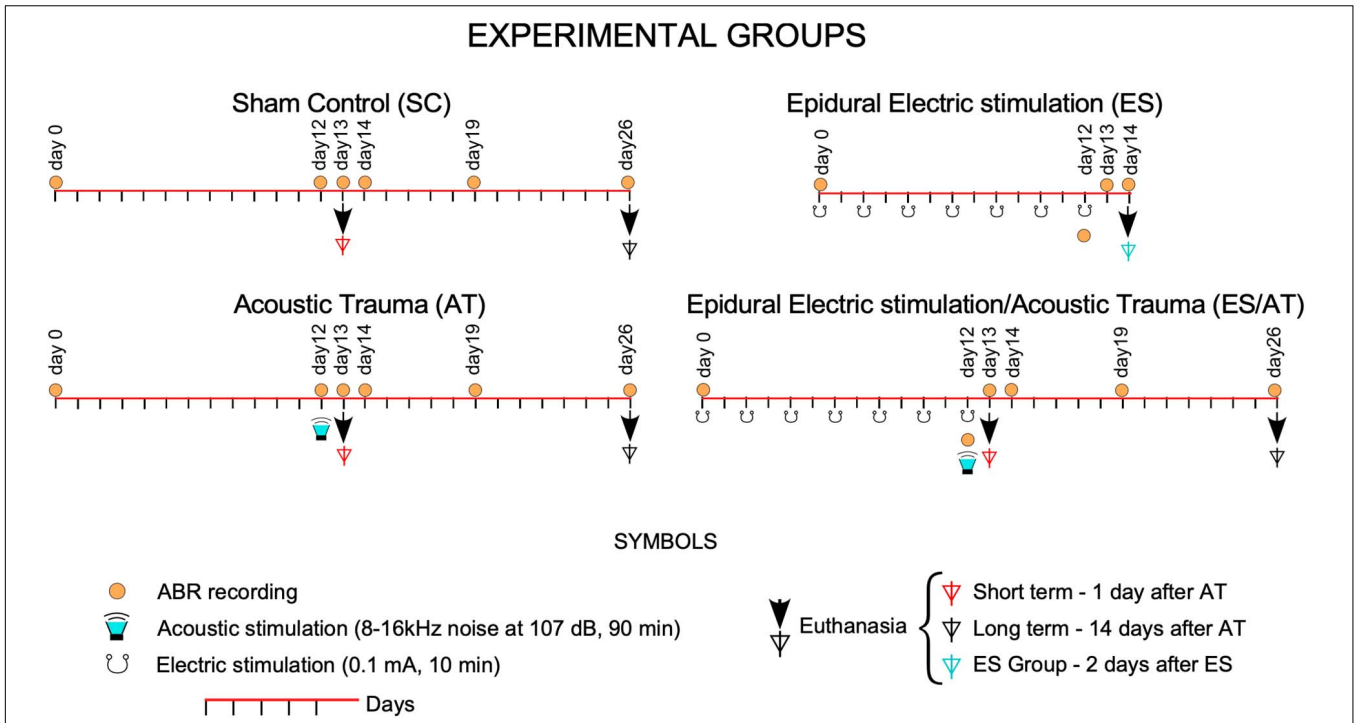

FIGURE 1 | Experimental groups: Timeline of intervention protocols. Over the line of time (red lines), day 0 labels the start of the protocols at 7 days after surgery. From day 0, AT and ES/AT groups underwent acoustic overstimulation at day 12 (green loudspeakers). In groups with electrical stimulation (ES and ES/AT), the multisession protocol was applied from day 0 to day 12. Symbols of ABR recordings (orange circles) and euthanasia (arrows and triangles) were over the curve of time (line in red) following the sequence of events in the protocols.

cranial surface. Following the Paxinos and Watson stereotaxic coordinates (Paxinos and Watson, 2005), four points were drawn on the surface of the bone delimiting the borders of the auditory area (for details, see Lamas et al., 2017). An approximately square window was carefully drilled on the bone surface until exposure of the surface of the dura mater. Cold saline (4°C) was dripped to avoid thermal cortical lesions. A 2.25 mm^2 silver electrode (anode) was gently encrusted into the trepans, and two screws (cathode) were implanted in the contralateral rostral-most side of the skull. After appropriately connecting the system, the electrodes and screws were fully covered by dental cement before any further intervention.

AC Epidural Stimulation

A 0.1-mA continuous current was delivered for 10 min per session through the epidural bone-attached electrode (anode) using an ISU 200 BIP isolation unit controlled by a CS-20 stimulator (Cibertec, Madrid, Spain). The stability of the voltage current was monitored along sessions. The electrical stimulation protocol was applied in awake animals for seven sessions on alternating days (days 0–12 of the protocol) (**Figure 1**). For more details, please see Colmenárez-Raga et al. (2019). To assess if the cortical damage after ES enables AC to drive corticofugal responses, serial sections of brain AC were immunostained for GAD 67 in rats from the ES group (please see below).

ABR Recordings

Recordings were performed under gas anesthesia using a real-time signal processing system [RZ6 Multi I/O Processor, Tucker-Davis Technologies (TDT), Alachua, Fl, United States]. The sound system outputs were calibrated before the recordings using a one-quarter-inch microphone (Brüel and Kjaer). Sound stimuli were 0.1-ms alternating polarity clicks, with a repetition rate of 21 clicks/s delivered in 10-dB ascending steps from 10 to 90 dB. The stimulation sessions were performed in an acoustically isolated chamber. The stimuli were delivered in a close field using a magnetic speaker (MF1 Multi-Field Magnetic Speaker TDT) connected to the ear through a 10-cm-long plastic tube. This approach resulted in a total delay of 1.4 ms in stimulus arrival at the tympanic membrane. ABRs were recorded by averaging 1000 EEG responses to 1000 click stimuli. Three subcutaneous needle electrodes were placed at the vertex and the two mastoids. Evoked potentials were amplified and digitized using a Medusa RA16PA preamplifier and a RA4LI head stage (TDT). Monaural ABRs were recorded from the vertex using the electrode on the mastoid ipsilateral to the click-stimulated ear, as the reference electrode. The needle in the mastoid contralateral to the stimulated ear served as the ground electrode. Monaural ABRs were sequentially recorded by click stimulation in the left and right ears. The placement of the recording electrodes was changed accordingly to record the signals from the side of the sound-stimulated ear. ABR recordings of both sides were analyzed separately. The final signal was filtered with a 500-Hz high-pass filter and a 3,000-Hz

low-pass filter (for more details on the ABR recording method, see Colmenárez-Raga et al., 2019). Wave II was first recorded in ABRs and then used to calculate thresholds (©MatLab R-2017 a). The ABR threshold was defined as the minimal sound intensity that evoked a significant voltage change (in a latency range between 1.4 and 5 ms) exceeding the mean ± 2 standard deviations of the voltage value of background activity during the first ms of the recording. The absolute wave latency was defined as time, in milliseconds, from the stimulus onset to the positive peak of the wave. The amplitudes of the ABR waveforms were measured as the peak-to-peak amplitude between the preceding negative trough to the subsequent positive peak of a given wave. In AT and ES/AT groups, ABRs were recorded before and after surgery, as well as 7 days after (day 0), right before AT (day 12) and 1 day (day 13), 2 days (day 14), 7 days (day 19), and 14 days (day 26) after AT (**Figure 1**).

Sound Stimulation for Acoustic Trauma (AT)

Awake rats were in a non-reverberant cage with non-parallel sides and exposed to a bandpass noise (8–16 kHz) of 107 dB for 90 min. Noise stimuli were generated digitally (RP2.1, TDT), filtered (RPVDS software), amplified (Audio Source AMP ONE/A), and calibrated inside the cage before each experiment using a one-half-inch microphone (Bruel and Kjäel Instruments, 4134) and a sound level meter with a fast Fourier transform (FFT) analyzer (Larson Davis 831).

Histology

Animals were deeply anesthetized with an intraperitoneal injection of 6% sodium pentobarbital (60 mg/kg BW) and perfused transcardially with 4% p-formaldehyde in a 0.1-M phosphate buffer (PB). Immediately, cochleae were perfused through the round window, dissected, postfixed for 2 h at room temperature, and decalcified in 8% EDTA for 12 days. Surface preparation membranes were extracted and then dissected into six pieces for whole-mounting processing of the cochlear epithelium. Immunostaining started with a blocking buffer (PBS with 5% normal horse serum and 0.3% Triton X-100) for 3 h, at room temperature, followed by a 2-day incubation at 37°C with the primary antibody, Goat Anti-Choline Acetyltransferase polyclonal antibody (AB144P; Merck Millipore, Temecula, CA, United States) at 1:100. After washing three times in TBS-Tx for 15 min, the dissected pieces were incubated with an anti-goat biotinylated secondary antibody (biotinylated anti-goat IgG H + L, BA-5000; Vector, Burlingame, CA, United States) at 1:200 for 24 h at room temperature. The pieces were then washed with TBS-Tx and incubated for 24 h in avidin/biotin–peroxidase (ABC complex, Vectastain Standard ABC Kit PK-4000; Vector, Burlingame, CA, United States) and further washed with TBS-Tx, followed by Tris–HCl, pH 8.0. They were then incubated in 3,3-diaminobenzidine tetrahydrochloride (DAB; D-9015; Sigma-Aldrich, St. Louis, MO, United States) with 0.006% H_2O_2 to visualize the peroxidase reaction. The pieces were finally dehydrated in graded alcohol solutions from 50 to 100%, followed by clearing in xylene and coverslipping.

To locate the area of stimulation in the cortex, brains were serially sectioned in the coronal plane into 40-μm sections and immunostained for IBA1 and GFAP, according to our previously published method (Colmenárez-Raga et al., 2019). Glial reaction, both for IBA-1 and GFAP, was delimited on the auditory cortices (data not shown) as previously described by our group. To analyze the state of preservation of the temporal auditory area, after multisession protocol, alternate serial sections of brains from the ES group were stained for Nissl and for GAD 67 monoclonal mouse antibody (Merck Millipore #MAB5406 clone 1G10.2 RRID: AB_2278725) diluted at 1:1,000 TBS 0.05 M + Triton-Tx 0.3% according to the protocol previously described in Pernia et al. (2020) (**Figure 2**).

Morphometry and Densitometry

For cochlear reconstruction, dissected immunostained segments of the organ of Corti (surface preparations) were photographed at × 5 objective and digitized using the Neurolucida software (NL-Vs 8.0, MicroBrightField®, Inc., Williston, VT, United States) under a Leica DMRX microscope equipped with a set of plan apochromatic objectives. Pictures from each slide were combined and ordered cochleotopically using as reference changes in the width of the organ of Corti and the thickness and density of the spiral bundle. Using this approach, a single final image of the whole cochlea was recomposed using the Canvas software (Canvas Draw 5 for Mac). After digital reconstruction, a line was drawn along the spiral bundle (SB) to calculate the cochlear length. These lines were measured using the Canvas perimeter tool. The length of the rat basilar membrane has been previously analyzed and estimated as 9.4 mm for Wistar Rats (Burda et al., 1988). In our samples, the mean perimeter of all cochleae measured was 8.08 mm (SD 0.79). According to previous measurements (Burda et al., 1988), our larger cochlear reconstruction was 9.4 mm in length. By using the "measure line" plugin of the ImageJ software program, provided by Eaton-Peabody Laboratories, the locations of several frequencies in the reconstructed cochlea were labeled for subsequent topographic cochleotopic analysis (see below). Once the frequencies were located in the cochlea, six pictures per cochlea (×40 objective) centered on 2.8, 8, 11.3, 16, 32, and 45.2 kHz were captured using the deep focus tool from Neurolucida 8.0 (MBF Bioscience, Williston, Vermont, United States). To obtain the resulting deep-focus image, five 1-μm images were Z-stacked using the first surface plane of sharp focus of immunoreactive terminals as a reference. The acquired images were processed for morphometric and densitometric analysis of ChAT immunostained buttons using the software ImageJ. Both densitometric and morphometry analyses were performed after separating (digital cutting) the entire OHC area in the pictures with the free hand tool of the ImageJ software program (**Figure 3**).

Morphometry

Size and Number of Terminals

Images (×40) with a digital resolution of 25 pixels/μm^2 were used for morphometry (**Figure 3A**). For segmentation of ChAT-ir cell buttons, thresholding operations were further

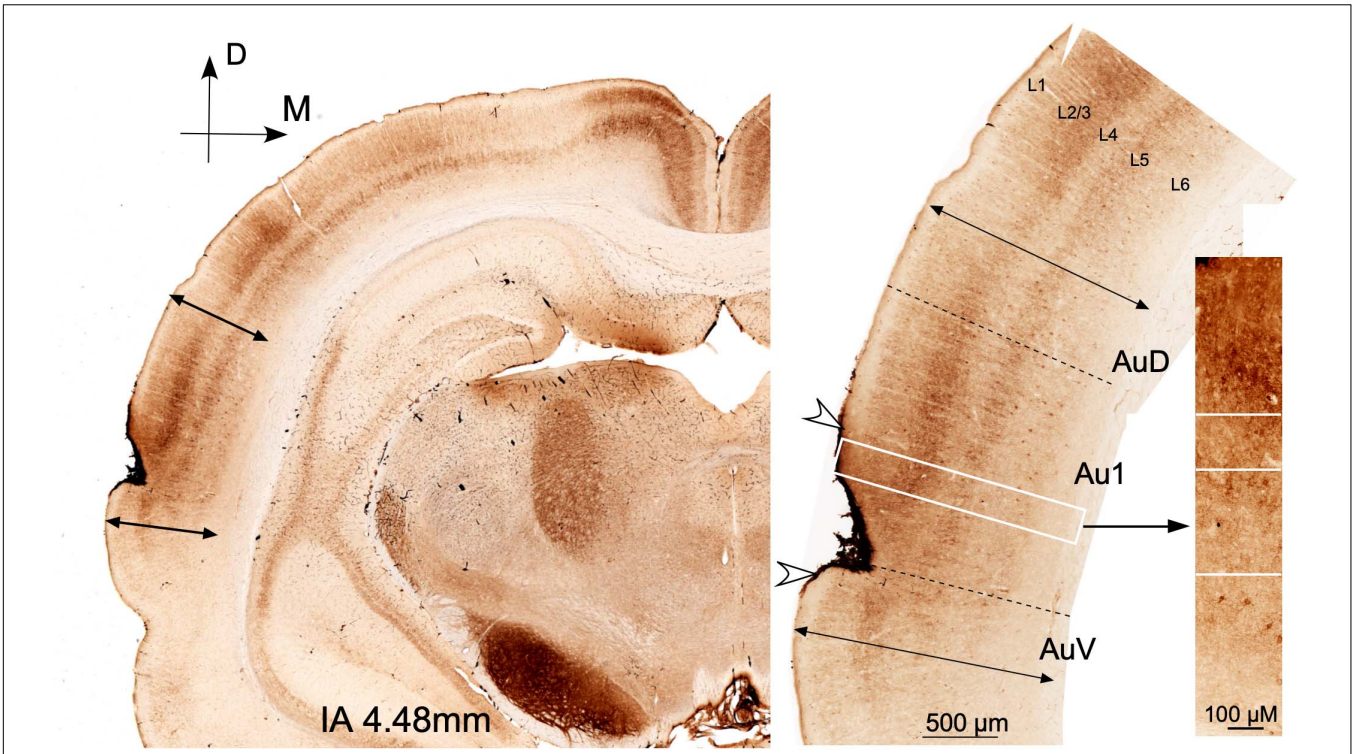

FIGURE 2 | Coronal sections at interaural 4.48 mm (Paxinos atlas coordinates) from a case from group ES euthanized at day 14 of the protocol (2 days after multisession epidural stimulation). GAD 67 dense reaction product in the surface of the cortex defines the position of the electrode. Arrowheads indicate the limits of the lesion in the coronal plane. The perimeter length of the reinforced area in the brain surface was 2.03 mm. Cytoarchitectural landmarks (double arrows) delimit the auditory cortex area. Note the well-defined immunoreactive layers and the well-preserved cytoarchitectural subdivisions in the non-damaged auditory area.

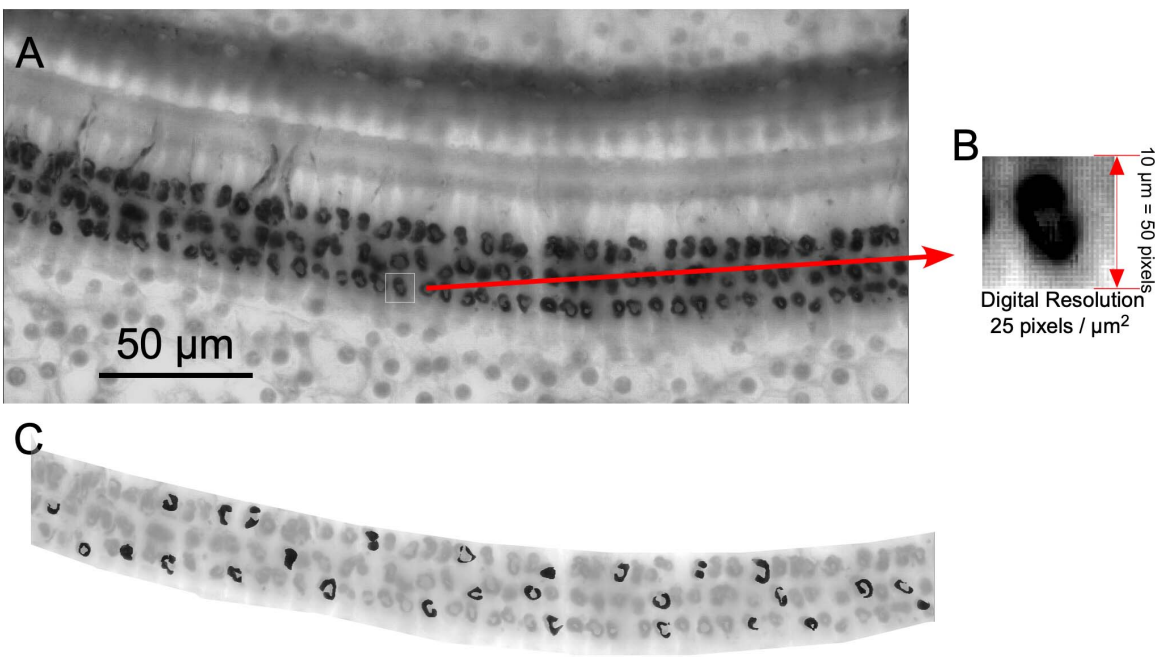

FIGURE 3 | Methodology used for densitometric analysis of immunoreactive buttons. **(A)** Example of a × 40 deep focus picture. **(B)** Digital resolution. **(C)** Thirty terminals per frame (highlighted in black) were manually taken by using freehand tool (ImageJ software).

applied with ImageJ. Thresholding was applied to all images, followed by automatic counting of all selected particles. Terminal immunoreactive buttons were segmented using density thresholding in the ImageJ software program. The number of segmented buttons was normalized to N/10,000 μm^2 surface area.

Densitometry

Before capturing, the illumination source of the microscope was adjusted using a stepped density filter (11 levels) (®EO Edmund industrial optics—ref 32599, Karlsruhe, Germany). In total, 30 buttons per ×40 image (equivalent to one frequency sample) were manually segmented using the ImageJ freehand selection tool (**Figure 3**). The density values of immunoreactive terminals were determined using the ImageJ software. The mean gray level of the neurons (a value between 0 and 255) was used as a measure of the button immunoreactivity to ChAT. We used the values of microscopic illumination determined using the density step filters (see above) to translate gray values into optical density (OD) values. In this paper, normalized gray OD levels were used instead of direct gray-level measures. The normalized gray levels were calculated by subtracting the mean of OD of the field (value of the entire OHC region) from the OD level of the immunoreactive buttons and by dividing the result by the standard deviation of the entire field.

Statistical Analysis

Statistical analysis was performed using the IBM® SPSS® software, version 25 (IBM Corp. and SPSS Inc., Chicago, IL, United States, RRID: SCR_002865). Differences in ABR thresholds values between different record times within each group were analyzed using the non-parametric Friedman test followed by Bonferroni *post hoc*. Comparisons between groups at each recording time were performed using the Mann–Whitney test. No significant differences were found when comparing recordings of the left and right ears of SC and stimulated animals.

For quantitative immunocytochemistry, one-way ANOVA followed by the Bonferroni and Games–Howell *post hoc* tests were used to assess differences between groups in OD, number/10,000 μm^2, and size of ChAT immunoreactive terminal buttons. Differences between groups by frequencies were assessed by two-way ANOVA. Spearman's rank and Pearson correlation coefficients were used to analyze correlations between size and OD measurements of immunoreactive terminals. Differences were considered significant at $p < 0.05$.

RESULTS

ABR Recordings

SC animals showed regular, constant 10-dB ABR thresholds in recordings at different timepoints of the protocol. In the AT group, the thresholds significantly increased at AT + 1 day (day 13) (30 ± 6.32 dB, $p < 0.01$) and AT + 2 days (day 14) (21.66 ± 4.08 dB, $p < 0.05$) (**Figure 4A**) and decreased at AT + 7 days (day 19) (13.33 ± 5.16 dB), albeit non-significantly, until reaching values similar to those of pretreated rats at AT + 14 days (day 26) (**Figure 4A**). In the ES/AT group, during the pretreatment period (before acoustic stimulation and after surgery, from 0 to 12 days), the animals received seven sessions of epidural electrical stimulation on alternating days (**Figure 4B**). After this sequence of cortical stimulation, and before AT, the mean thresholds increased to 33.33 dB (SD 5.16) (**Figure 4B**). However, no significant differences in mean thresholds were found at AT + 1 day (day 13) (**Figure 4B**). The comparison of the thresholds at the same stages of the protocol AT + 1 day (day 13) and AT + 2 days (day 14) between the AT and the ES/AT groups shows that the means of the AT group are significantly higher than those of the ES/AT group (**Figure 4C**). In the ES group, the thresholds increased after the last session of epidural stimulation (day 12), as shown in the ABR recordings. The mean threshold values reached normal levels at ES + 1 day (day 13) (**Figure 4D**).

Brain Cortex Preservation After Multisession Stimulation

The localization and extension of the electrical stimulated area in the brain cortex was analyzed using glial immunocytochemical markers (GFAP and IBA-1) and Nissl staining in alternate sections, following the approach previously applied by our group (Colmenárez-Raga et al., 2019). In addition, to test the state of preservation of the cortical microcircuitry in the ES animal group, serial sections were stained for GAD 67 (**Figure 2**). As observed in our previous study, all areas of glial reaction highlighted by glial markers were restricted to the auditory temporal area. Furthermore, the more superficial layers were affected in varying degrees, depending on how the electrode is encrusted into the skull [data not shown; please refer to Colmenárez-Raga et al. (2019) for further details].

Denser GAD 67 immunoreactivity in the dura and superficial layers of the cortex makes it possible to define the extension of the damaged areas (**Figure 2**). Immunoreactive GAD neurons and terminal fields are present virtually throughout the auditory cortices despite a ribbon under the dura (**Figure 2**). Around the area of contact of the electrode, the cytoarchitecture and layering of the auditory temporal area can be easily differentiated through cases, thus indicating that cortical microcircuits beyond the damaged region are well preserved (**Figure 2**).

Anatomy

General Features From Sham Controls

Under the microscope, immunostained preparations showed thick, myelinated fascicles of positive efferent fibers across the spiral limbus of Huschke and along the floor of the tunnel of Corti. Fascicles penetrated in, ramified at, and meandered around the deeper section of supporting phalangeal cells. After a short ascent, the fibers ended in terminals and innervated the basal pole of the OHC. A dense dark-brown reaction product sharply defined the size and shape of efferent terminals along the cochlea (not shown). Differences in size, density, and number of immunoreactive buttons were observed along the frequency range (**Figure 5A**). When comparing buttons between cochleotopic regions, the higher the frequency, the more regular the distribution

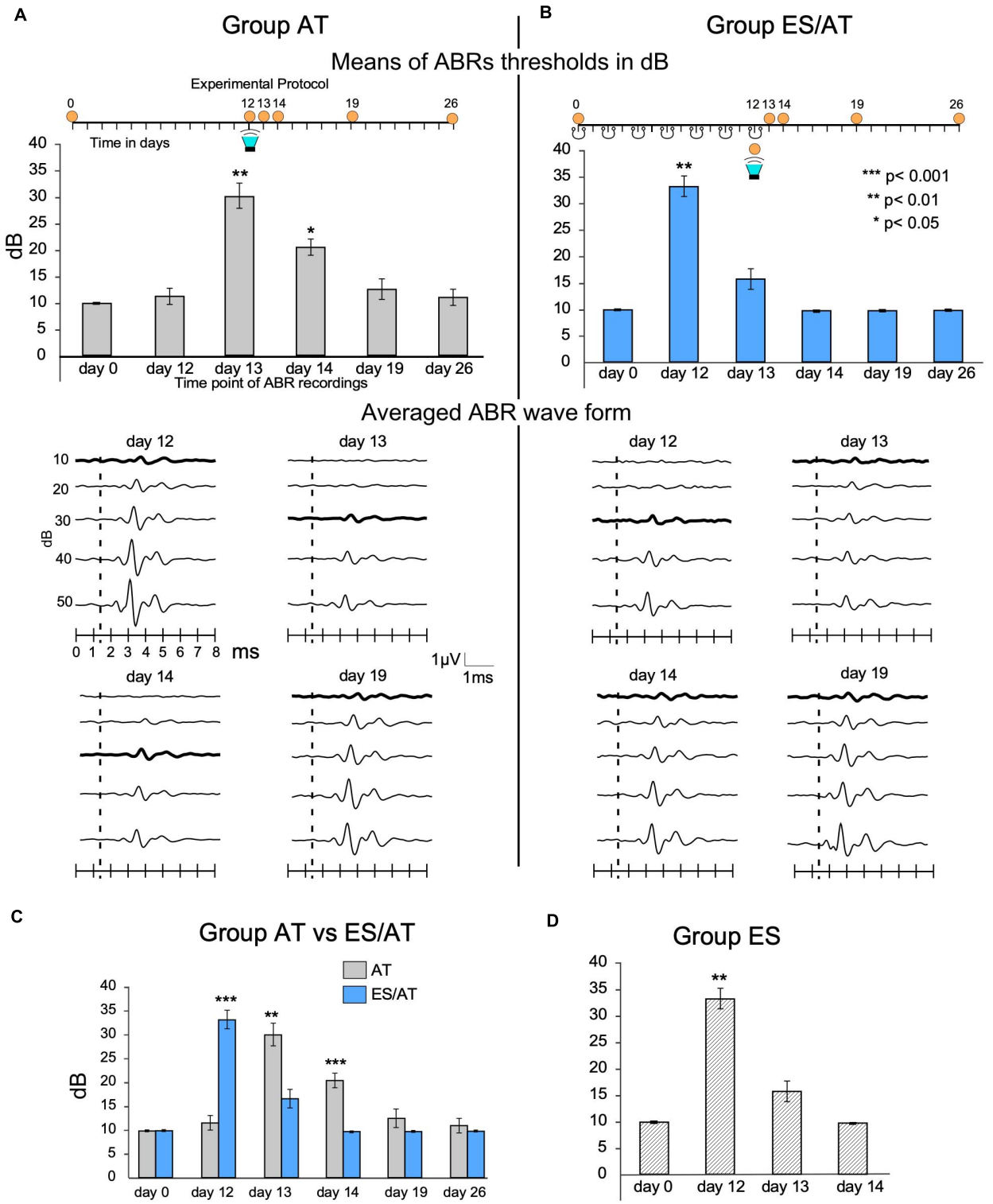

FIGURE 4 | Thresholds and waveform analysis of ABRs. **(A)** The AT group shows a significant threshold shift at AT + 1 day (day 13) and AT + 2 days (day 14). **(B)** The ES/AT group shows a significant threshold shift before, but not after, overstimulation (the green loudspeaker labels acoustic stimulation). Waveform analysis is shown in the middle of the panel. Lines represent averaged waves from all animals of each group. The thickest lines label averaged thresholds. **(C)** Statistical comparison of the mean thresholds of groups AT vs. ES/AT. Note significant differences in threshold shifts after AT at day 13, the day after acoustic stimulation. Please also note that, in the days after the recordings, the thresholds recover later in the AT group than in the ES/AT group. **(D)** ES group. Multisession epidural stimulation induces threshold shifts immediately after the last stimulation on day 12, which recovers 1 day later (day 13).

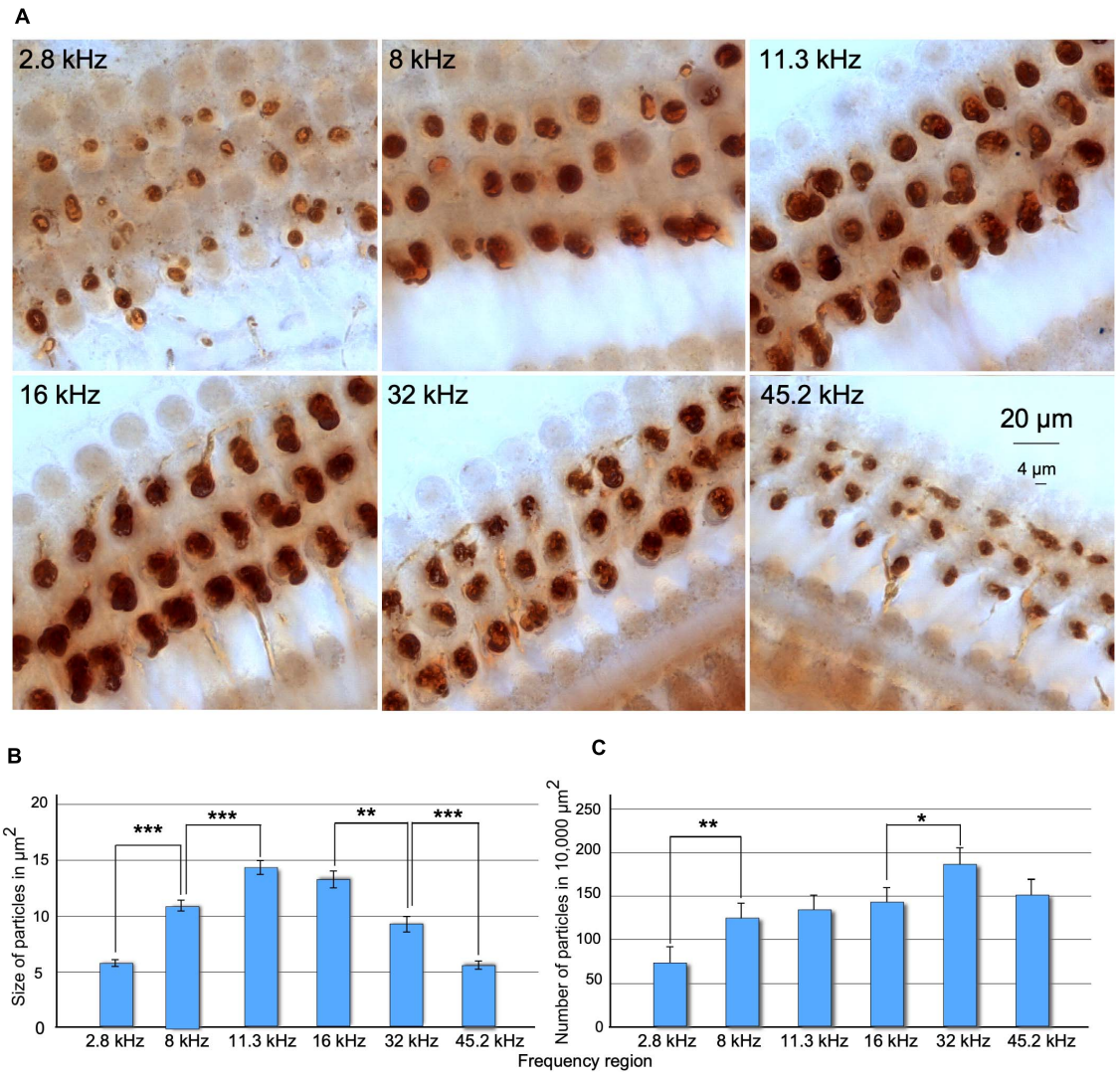

FIGURE 5 | (A) Comparison of changes in size and number of ChAT-immunostained terminal buttons along frequency regions of the cochlea in SCs. Images were acquired using a deep focus tool. **(B)** Statistical analysis of size of buttons upon density threshold segmentation. Note that the size of the terminals increases and decreases along the tonotopic axis of the cochlea. Significant differences were found when comparing each frequency region with the adjacent, except between 11.3 and 16 kHz regions. **(C)** Number of terminals measured by density threshold segmentation normalized to 10,000 μm^2. The lower significant number of the terminals was found at 2.8-kHz areas and the highest at 32-kHz areas. After comparing adjacent frequency regions, significant differences were found between 2.8 and 8 kHz and 16 and 45 kHz (*$p > 0.05$, **$p < 0.01$, ***$p < 0.001$).

in rows and the shape of immunoreactive buttons would be (**Figure 5A**). Values in the size of the buttons increased gradually in mid-frequency regions and decreased in high-frequency regions (**Figure 5B**). Overall, normalized numbers of segmented particles indicate a gradual increase, along the cochlear axis, from low- to high-frequency regions (**Figure 5C**).

Inflammatory Response

In both SCs and stimulated cochlea (AT and ES/AT), free cells were more frequently found in the tunnel of Corti. These cells were irregular in shape and variable in size (from 5 to 20 µm) (**Figure 6**). The largest cells (10–20 µm), which showed intense ChAT immunoreactivity, were spherical and contained filopodia and pseudopodia, features which identify them as macrophages (**Figure 6** arrowheads). The smallest cells (about 5–10 µm) were not immunopositive but also had filopodia and were thus compatible with monocytes (**Figures 6A,B,F** white arrows). A few of these cells were found in control cochleae. For this reason, we considered them as a casual finding at the beginning of the microscopic observation. Some of these free cells have been found in our material associated with the loss of efferent synapses and the presence of debris, which suggest an active digestion of terminal buttons (**Figures 6B–D**). In surface preparations of stimulated cases (AT and ES/AT), the number of macrophages was extremely variable

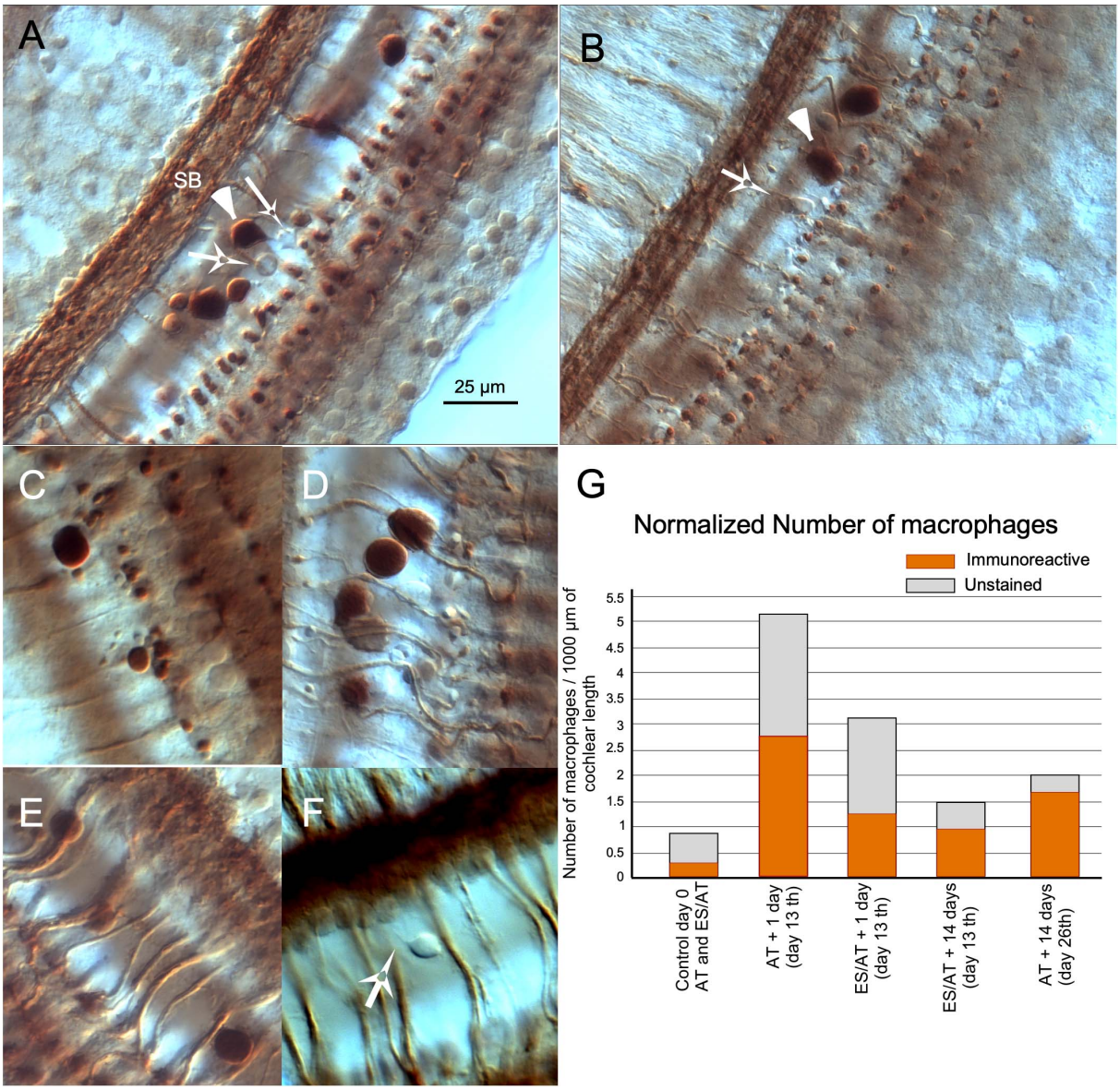

FIGURE 6 | Inflammatory reaction in the cochlea after acoustic trauma. **(A)** Panoramic view of the organ of Corti in surface preparation. Large immunoreactive inflammatory cells (arrowheads) are interspersed with other smaller non-immunoreactive (arrows). **(B)** Inflammatory cells, presumably macrophages, migrating to the OHC area. **(C,D)** Details of close contacts of immune cells with immunoreactive terminals. Cell debris is associated with the loss of terminals. **(E)** Immune cells climbing along efferent fibers. **(F)** Presumably monocytes with filopodia, most frequently observed in the floor of the tunnel. Differential interference contrast microscopy (Nomarski) (×63 PLAN APO Leica/NA 1.32-0.6 immersion oil). **(G)** Quantification of the immune response by counting separately the number of immunoreactive and unstained macrophages. The number of inflammatory cells is higher after AT in the AT group and decreases at 14 days (day 26th). SB spiral bundle.

with a random distribution along the cochlea. After separately counting immunoreactive and non-immunoreactive cells and normalizing the values (per 1,000 µm of cochlear length), a higher number of cells were tallied AT + 1 day (day 13) than ES/AT + 1 day (**Figure 6G**). In both experimental conditions (AT and ES/AT), the values of number of reactive and non-reactive inflammatory cells decreased at AT + 14 days (day 26) (**Figure 6G**).

Cholinergic Olivocochlear Terminal Buttons. Quantitative Immunocytochemistry Morphometry

After counting immunoreactive buttons by density threshold segmentation, a statistically significant loss of terminals was not detected in any experimental groups when comparing with SC (data not shown). A significant decrease in the size of terminals

was detected only in group AT + 1 day (day 13) when analyzing the cochlea as a whole ($F = 3.639$; $p < 0.01$) (**Figure 7A**). However, in both experimental groups (AT and ES/AT) a non-significant decrease was also observed in ES/AT + 1 day (day 13), with a recovery of the values at 14 days after AT (day 26) (**Figure 7A**).

Densitometry

Analysis of the whole cochlea showed a significant decrease in normalized OD values when comparing SC with both experimental conditions (AT and ES/AT) at 1 day after acoustic trauma (day 13) ($F = 3.548$; $p < 0.001$) (**Figure 7B**). In addition, OD values significantly rebounded in both experimental conditions (AT or ES/AT) at 14 days after acoustic trauma (day 26), although the ES/AT values were significantly higher (**Figure 7B**).

Analysis by Frequency Regions

Six frequency regions were selected along the tonotopical axis per cochlea (see "Materials and Methods" section) and analyzed to assess changes in size and OD of immunoreactive terminals (squares in cochlear reconstruction in **Figure 8A**). In the SC group, the values of the size of buttons gradually increase and then decrease across frequencies, showing a Gaussian distribution, peaking at 11.3 kHz (**Figure 8B** top). In comparison, the size of terminal buttons in AT and ES/AT animals decreased at AT + 1 day (day 13), mostly in the middle-frequency regions (please see **Figure 8B**, red lines in AT + 1 day and ES/AT + 1 day). Changes in the shape of the lines connecting mean size values allows a better understanding of the evolution of the parameters along frequency areas (envelopes—red lines in **Figure 8B**). Statistical comparison of the values for each frequency region with the SC group showed a significant decrease in the 11.3-kHz region at AT + 1 day (day 13) (**Figure 8B**, arrow). In addition, at 14 days after ES (day 26), the size for terminals increases again for the middle frequencies in both groups (AT and ES/AT) recovering the Gaussian distribution observed in SC (**Figure 8B** right). Interestingly, the largest terminals appear now in the 16-kHz region, in contrast to peak in the 11.3-kHz region of the SCs (**Figure 8B**—compare bars highlighted in yellow).

The distribution of OD values in the SC group showed no significant differences across frequency regions (**Figure 8C**—top). However, a significant decrease in values was evident in AT + 1 day (day 13), in the 8- and 32-kHz regions (**Figure 8C**—arrows). Such decrement was not found in the ES/AT group (**Figure 8C** bottom). On the other hand, at day 26 of the protocol, the OD is increased compared to the sham controls in group ES/AT in the 2.8- and 45.2-kHz regions (**Figure 8C** bottom). After correlation test analysis of both parameters (size and OD), all experimental groups, except ES/AT + 1 day (day 13), showed a significant correlation (**Table 1**). We suggest that this finding speaks in favor of a better-preserved cochleotopy (more similar from controls) 1 day after ES.

DISCUSSION

In this study, we have shown that multiple sessions of electrical activation of the AC before sound overstimulation preserve hearing thresholds and curtail the inflammatory response in the cochlea without a significant loss of terminals. Furthermore, sound overstimulation significantly reduces the size of the immunoreactive cholinergic buttons 1 day after acoustic overstimulation (day 13) in the AT group, but not after cortical electric activation in the ES/AT group as shown by ChAT quantitative immunocytochemistry in cochlear surface preparations. The OD values of the ChAT immunoreaction products decrease in both experimental groups (AT and ES/AT) at 1 day after sound overstimulation (day 13). The values of both parameters (size and OD) recover at 14 days after acoustic overstimulation (day 26) despite OD means being significantly higher in ES/AT than in AT.

The analysis of the normalized measurements of immunoreactive buttons by frequency region shows statistically significant decreases only in AT + 1 day (day 13), for both size (at 11.3 kHz) and OD (at 8 and 32 kHz). Correlation test analysis for both parameters (size and OD) shows no significance only in group ES/AT + 1 day (day 13).

Animal Model

Cortical Effects of Epidural Stimulation

Changes in GAD immunoreactivity after cortical damage were previously analyzed by our group in a model of restricted ablation of the AC, showing that this marker for GABA neurons allows to define the limits of the lesion as well as the cortical cytoarchitectural subdivisions (Lamas et al., 2013). Inhibitory microcircuits (GAD-GABA) are crucial for neuronal network regulation (Kawaguchi, 2017) and indirectly reflect, if well preserved, potential effectiveness for driving responses of the brain cortex. Both present results (**Figure 2**) and our previous analysis of the effects of cortical multisession electric stimulation (Colmenárez-Raga et al., 2019) have shown that injuries in the cortex after electrode activation are restricted to relatively small areas of the auditory temporal area. Moreover, descending pathway activation of ACs can be ensured in our material, since the deeper cortical layers (layers 5 and 6, where corticofugal neurons are located) are not significantly affected (**Figure 2**). The size and shape of the damaged temporal cortex, as shown in reconstructions from our previous publication (please see Figure 9 in Colmenárez-Raga et al., 2019) and the well-preserved GAD cytoarchitecture (present results, **Figure 2**), indicate that auditory cortices remain functional after the protocol of ES. Since stimulation is unilateral in our model, excitatory callosal connections may also contribute to drive the corticofugal neurons of the contralateral side. Although residual plasticity effects cannot be fully ruled out in our experimental approach, marked increases and decreases in ABR threshold shifts also indicate a dynamic active feedback regulation of cortical neuronal networks over time. Notwithstanding anatomical analysis of AC preservation after stimulation, minimal lesions should be considered out of safety limits.

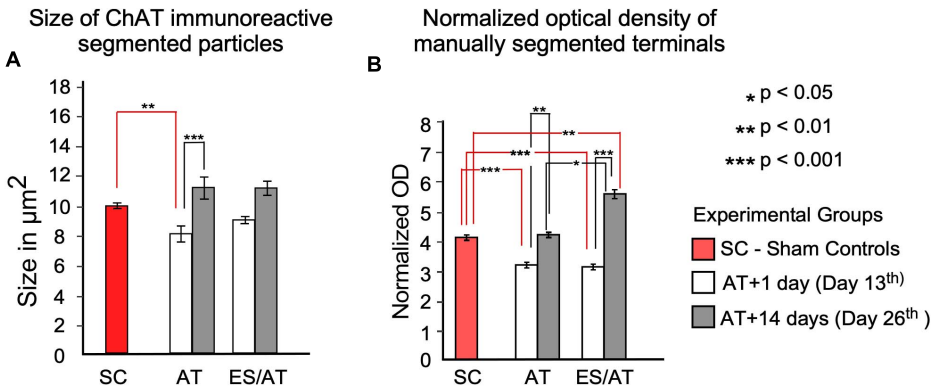

FIGURE 7 | Area and OD statistical analysis of ChAT-immunoreactive terminals through the cochlea. **(A)** Size of the particles. Differences were significant between the SC group and AT + 1 day (day 13) (red lines indicate this comparison). After comparing day 1 (day 13) and 14 (day 26) days after overstimulation, significant differences were identified only in the AT group (black lines). No significant differences were found after any comparison in the ES/AT group. **(B)** OD analysis of efferent terminals. Values are significantly lower at 1 day after overstimulation (day 13) under both experimental conditions (with or without ES). A rebound in values, significantly higher in the ES/AT group, was found at 14 days after overstimulation (day 26). $*p < 0.05$, $**p < 0.01$, $***p < 0.001$.

Sound Overstimulation

Our experimental approach to overstimulation was quite similar to the one published by Lobarinas et al. (2017). These authors reported (after subjecting rats to 2 h of band-pass noise of 8–16 kHz at 106 dB) threshold shifts ranging from 20 to 25 dB, approximately 1 day after sound overstimulation and a long-term full recovery, as shown in our animal model. Furthermore, in this paper, tonal ABRs show threshold elevations with a linear increase in the values from lower to higher frequencies and with a decrease in distortion product otoacoustic emissions (DPOAE) amplitudes (Lobarinas et al., 2017).

In a similar overstimulation protocol in mice, threshold shifts 1 and 2 days after overstimulation recovered at 8 weeks after applying a band noise of 8–16 kHz (100 dB, for 2 h in free field) (Kujawa and Liberman, 2009). These authors also reported an acute loss of synaptic ribbons in hair cells, which may have functionally silenced neurotransmission despite the complete recovery of hair cell function. Such an alteration, known as hidden hearing loss (HHL), supports an underlying long-term alteration of neurotransmission to spiral ganglion dendrites, after which auditory thresholds resume normal values. However, cochlear nerve responses depend not only on glutamatergic neurotransmission of IHCs but also on efferent cholinergic neurotransmission of outer hair cells, which indeed regulates micromechanically its responses (Malmierca and Merchán, 2004). Our results show changes in efferent neurotransmission with normal thresholds at 14 days after AT (day 26) (**Figures 7**, **8**). Following this line of thinking, we suggest that our protocol, or other similar ones, should be explored in the future at longer survival times together with a combined evaluation of afferent and efferent neurotransmission.

Hearing Sensitivity (ABRs)

Multiple sessions of ES before sound overstimulation induce hearing suppression, as shown by elevated thresholds after ABR recordings (**Figures 4B,D**), thus confirming our previously reported results using a similar stimulation protocol (Colmenárez-Raga et al., 2019). Due to the excitatory character of the corticofugal descending pathway (Feliciano and Potashner, 2002), anodal stimulation of the AC can drive the direct (cortico-olivary) and indirect (*via* inferior colliculus IC) corticofugal pathway (Horváth et al., 2003), which ultimately activates MOC olivary neurons. From an anatomical point of view, direct connections from infragranular layers of the AC and from the IC to the olivary complex support a descending corticofugal activation of VNTB–MOC neurons (Spangler et al., 1987; Vetter et al., 1993; Malmierca et al., 1996; Saldaña et al., 1996; Weedman and Ryugo, 1996; Winer et al., 1998; Schofield and Cant, 1999; Thompson and Schofield, 2000; Senatorov and Hu, 2002; Doucet et al., 2003; Warr and Boche, 2003; Bajo and Moore, 2005; Bajo et al., 2007; Malmierca and Ryugo, 2011; Mellott et al., 2014; Straka et al., 2015). Thus, in our protocol, persistent threshold shifts, shown after the end of cortical electrical stimulation (day 13) (**Figures 4B,D**), can be explained by a sustained activation of synaptic plasticity machinery in cortical networks. Such cortical activation has also been demonstrated after daily anodal transcranial direct current stimulation, which induces a persistent neural excitation and overexpression of plasticity-associated genes in the sensorimotor cortex (Kim et al., 2017). Moreover, after AC restricted ablation in the rat, the AC is able to trigger plasticity in the organ of Corti, inducing stable and long-term changes in the expression of motor proteins (Prestin and ß Actin), as previously shown by our group (Lamas et al., 2013, 2014). Considering the roles of the efferent system in hearing (Guinan, 2010), threshold elevation, after ES, can be explained by sustained and persistent cortico-olivary activation followed by MOC and/or inner ear plastic responses. Therefore, the effect of ES in decreasing hearing sensitivity (hearing suppression), before sound overstimulation (**Figures 4B,D**), is the most plausible explanation for differences in changes in threshold shifts over time between the AT and the ES/AT experimental groups. In addition, in the ES/AT group, the threshold shift occurred before AT, with normal thresholds at 1 day after AT (day 13); hence, changes in immunoreactivity parameters are

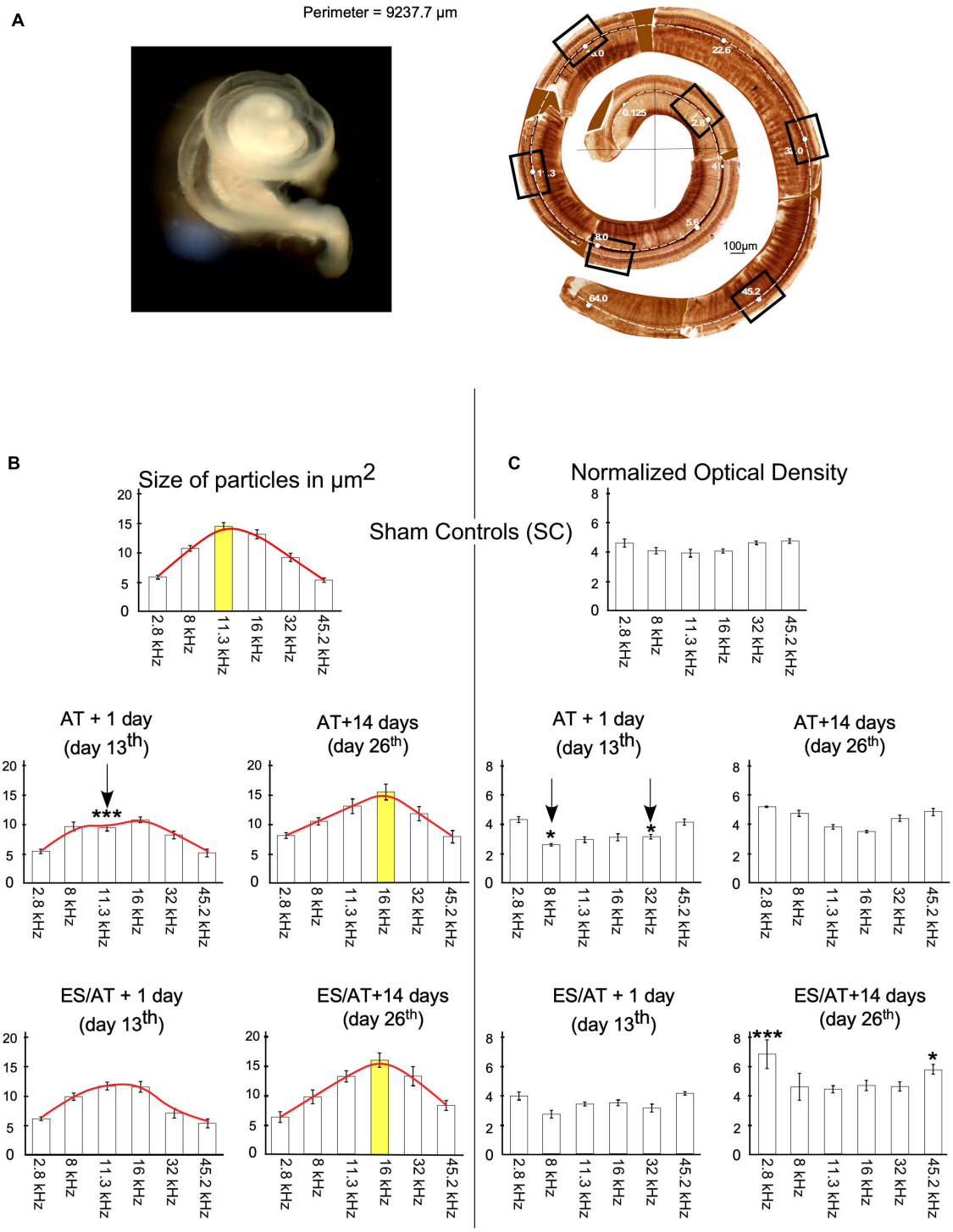

FIGURE 8 | (A) Image on the right shows membranous labyrinth of the cochlea after dissection. On the right, cochlear reconstruction was performed using digitally glued images from immunostained Organ of Corti surface preparation. Total length of the cochlea in this example was 9.237.7 µm, which ensures a complete extraction and accurately allows to define frequency regions. Squares delimit frames of the photographs taken to analyze a significant sample of six frequency regions. **(B)** Quantitative analysis of size by frequency region. Each bar in the graphs corresponds to one frequency region. Red lines on the top of the bars highlight the evolution of means of size values along cochleotopic regions. Note the changes in shape of the enveloping red lines in relation to SC at 14 days (day 26). Asterisks indicate significant differences between the experimental groups and SC for each frequency. Yellow bars indicate frequencies of higher values in SC, AT, and ES/AT groups at 14 days after AT (day 26). Arrows indicate significant decreases in values in the 11.3-kHz region at AT + 1 day (day 13). **(C)** Significant decrease in OD values at AT + 1 day (day 13) in the 8- and 32-kHz regions and increases in 2.8 and 45.2 kHz, in the ES/AT group.

TABLE 1 | Correlation test.

Animal groups	Means/SD of OD	Means/SD of area	Coefficient	*p*-value
Control	4.30/1.25	10.00/4.63	−0.123	$p > 0.05$
AT + 1 day	3.37/0.84	8.14/2.51	−0.584	$p < 0{,}001$
EE/AT + 1 day	3,26/0,79	9,11/3,11	−0.269	$p > 0{,}05$
AT + 14 days	4,34/0,69	11,36/3,73	−0.578	$p < 0.001$
EE/AT + 14 days	5.14/1.90	11.33/4.45	−0.391	$p < 0.001$

most likely primarily related to cortical stimulation, rather than to AT, under this experimental condition, as discussed below. Synaptic plasticity activation has been demonstrated by whole-cell patch-clamp recordings, showing that layer 5 neurons can respond through long-term potentiation (LTP) or long-term depression (LTD), after layer 6 stimulation in AC slices (Kotak et al., 2007). Also, tDCS stimulation increases cortical neuronal metaplasticity in AC neuronal networks (LTP or LTD) (Nitsche et al., 2008; Zhang, 2013). Moreover, after repetitive stimulation of the cortex, neurons develop a sustained increase in firing rate for hours (Nitsche and Paulus, 2000). In our recordings, presumable increases in the excitability and firing rate of corticofugal neurons of layers 5 or 6 after repetitive anodal stimulation can be related to MOC neuron activation. Threshold shifts induced by overstimulation depend primarily on reflex arc activation in the low auditory pathway (De Venecia et al., 2005). However, after electric stimulation of the temporal area and acoustic overstimulation, both feedbacks (reflex arc and cortico-olivary) involved in cochlear amplification work in combination. Therefore, the increase in thresholds after AT may be driven by persistent metaplasticity (LTP) of the epidurally stimulated cortex. Indeed, the recovery of thresholds at ES/AT + 1 day (day 13) indicates a compensation net induced by an overactivated corticofugal pathway acting on MOC olivary neurons.

Technical Limitations

The click stimulus used to record the ABRs is a broadband stimulus covering a wide range of low frequencies (<10 kHz). Although this click stimulus can be used to successfully measure threshold shifts in rats (see **Figure 4**), it may not show the potential contribution of high frequencies to transient threshold shifts. Accordingly, future studies using tonal ABRs may shed light on putative differences in threshold at low and high frequencies.

Quantitative Immunocytochemistry

ACh is synthetized in the soma and terminals of the neurons by the combined choline and acetyl CoA reaction catalyzed by ChAT. ACh is delivered into the synaptic clefts and coupled by receptors, and the remaining neurotransmitter is hydrolyzed by the enzyme acetylcholinesterase (AChE); concurrently, choline reuptake in the terminals enables its coupling with acetyl CoA (Simon and Kuhar, 1976; Kuhar, 1979; Matsuo et al., 2011). Consistently, neurotransmitter storage in terminals depends on the rate of synaptic delivery of ACh, on balanced *de novo* synthesis, and on neurotransmitter recycling. Our measurements of the size and OD of the buttons, which quantitatively indicate the amount of reaction product, and ultimately the rate of ACh synthesis, reflect the state of synaptic neurotransmission. Thus, decreases in the size and OD of efferent terminals shown by us reflect synaptic depletion, after sound overactivation. After sound overstimulation, AChE (the enzyme involved in ACh recycling) immunoreactivity in the organ of Corti decreases in guinea pigs (Mounier-Kuhn and Haguenauer, 1967) and chinchilla (Kokko-Cunningham and Ades, 1976) according with our results. In our samples, the size of the terminals significantly decreased in AT + 1 day (day 13) but not in ES/AT + 1 day (day 13). Thus, the differential values of terminal button quantitative analysis, assessed in ES/AT, can be related to a mitigated noise effect induced by hearing suppression. The significant rebound of OD values in the ES/AT group 14 days after AT (day 26) suggests a recovery in ACh synthesis after depletion by acoustic trauma. However, significant differences in OD values in the ES/AT group at 14 days (day 26) after AT (**Figure 8C** bottom), which were not observed in the AT + 14 days (day 26) group, may also be related to the activation of long-term cortical plasticity by electrical stimulation. In the ES/AT group at 1 day and 14 days after AT, quantitative analysis shows that changes in values of quantitative immunocytochemistry result in combination with normal thresholds (**Figures 4**, **7**). These results suggest that temporal windows for stabilization of efferent neurotransmission do not match hearing sensitivity recovery. Future analysis of correlation of efferent and afferent system alterations at long term will be needed to shed light on this problem.

Analysis by Frequency Regions

According to a previous ultrastructural analysis of ChAT immunoreactivity in the organ of Corti, MOC terminals in the OHC are compact and densely filled with reaction products (Eybalin and Pujol, 1987), as shown in our light microscopy images. Consequently, fulfilled terminals with a homogeneous immunoreactive product, shown in our material, ensure that our measurements detect accurately the actual size of the buttons. Synaptic size is affected by multiple molecular mechanisms, some of which depend on dynamic synaptic activation, whereas others remain unaffected (Lund and Lund, 1976; Pierce and Milner, 2001; Stanic et al., 2003; De Paola et al., 2006; Pasantes-Morales and Tuz, 2006; Grillo et al., 2013; Petrof and Sherman, 2013; Statman et al., 2014; Pasaoglu and Schikorski, 2016; Sammons et al., 2018). Cochleotopic analysis of the size of immunoreactive terminals in our control animals shows a progressive increase from low-to-medium-frequency areas and a gradual decrease to high-frequency ones (**Figure 8B**). The bell distribution of size values in the SC group is replaced by a more homogeneous distribution at AT + 1 day (day 13) and ES/AT + 1 day (day 13) due to the decrease in the size of the buttons at the middle range of frequencies (8–16 kHz). These results suggest a more intense effect of frequencies at the noise band (8–16 kHz) used for sound stimulation (**Figure 8B**). Unlike the size of the terminals, the OD values, as assessed by frequency region, show a flat distribution in the SC group, with slightly higher values at the ends of the frequency range (**Figure 8C**). In fact, OD measures the amount of ChAT in synaptic efferent buttons and, indirectly, the rate of ACh synthesis. A significant decrease in OD values of AT + 1 day (day 13) in the frequency regions

of 8 and 32 kHz (not shown in the electrical stimulated group) may reflect unrecovered neurotransmitter synthesis in mid-frequency regions after depletion of terminals by overactivation (**Figure 8C**). Tonotopical analysis of ABRs in rats, applying a similar TTS protocol (107 dB—frequency band of 8–32 kHz—90 min of sound exposure) and with a similar timeline (1 day and 2 weeks), has shown that, although thresholds returned to baseline, wave 1 amplitudes at 16, 24, and 32 kHz failed to return to control levels (Lobarinas et al., 2017). Significant changes in size and OD have been shown in our stimulated groups in similar ranges of frequencies at 1 day post stimulus but not at 14 days (**Figures 8B,C**); however, recovery was observed at 14 days after exposure. This apparent discrepancy in tonotopic effects of AT between our anatomical results and those of Lobarinas et al. (2017) can be related to a delayed recovery of ACh synthesis with respect to wave-amplitude thresholds. No significant changes in size or OD have been shown in the ES/AT + 1 day (day 13th) group across frequency regions. Conversely, significant differences were shown in both parameters at AT + 1 day (day 13th) (vertical arrows in **Figures 8B,C**). It seems relevant to remark that all experimental groups, except ES/AT + 1 day (day 13th), showed a significant correlation, after correlation test analysis, which suggests better-preserved cochleotopy (more similar from SC) induced by electric stimulation of the temporal cortex (**Table 1**).

Inflammatory Response

Immune responses, primarily involving monocytes and macrophages in the cochlea, have been shown after sound overstimulation (Fredelius and Rask-Andersen, 1990; Hirose et al., 2005; Wood and Zuo, 2017; Frye et al., 2019; He et al., 2020, among others).

The cochlear immune response includes resident cells, which can actuate by humoral liberation of inflammatory mediators (i.e., supporting cells and lateral wall fibrocytes, among others) (Cai et al., 2013) and mobile cells (macrophages). Such cochlear cleaners are located in the basilar membrane, as silent monocytes, which migrate to the sensory epithelium after activation by cochlear damage (Fredelius and Rask-Andersen, 1990; Frye et al., 2017). Our microscopic observations show ChAT-immunoreactive cells, variable in size (5–20 μm in diameter) and shape (irregular or globular), usually with filopodia, mainly located in the tunnel of Corti, which can be anatomically identifiable as macrophages (**Figure 6**). These cells have been closely related to areas of cell debris and loss of immunoreactive buttons (**Figures 6A–D**). Collateral pruning by microglial cells is currently considered a physiological mechanism of regulation of network connectivity and plasticity. Accordingly, microglial amputation of buttons occurs in neurological diseases, such as Parkinson's disease, Alzheimer's disease (Hong et al., 2016), epilepsy (Andoh et al., 2019), or schizophrenia (Sellgren et al., 2019). Whether or not macrophages, in our animal model, effectively or extensively participate in the remodeling of efferent terminal fields in damaged cochlea remains unknown, but this is undoubtedly an interesting question, which merits further research in the near future.

Considering that some ChAT immunoreactivity was detected interstitially, cochlear macrophagic cells may also be involved in actively removing the enzyme from the perilymph. However, dendritic macrophages physiologically express choline acetyltransferase (ChAT), muscarinic and nicotinic acetylcholine (ACh) receptors, and acetylcholinesterase (AChE) (Fujii et al., 2017). In principle, their potential constitutive molecular profile may also explain their immunoreactivity. Yet, small cells remained unstained, and the immunoreaction was mainly detected in the larger cells. This finding supports the hypothesis that reactive macrophages may act as cleaners of ChAT, after neurotransmitter depletion in the efferent terminals. Our normalized cell counts showed that the number of macrophages is higher in AT + 1 day (day 13) than in ES/AT + 1 day (day 13) (**Figure 6G**), which suggests that the immunoreaction (and presumably the cochlear damage) is lower after cortical electrical stimulation.

CONCLUDING REMARKS AND CLINICAL IMPLICATIONS

In this paper, we have shown that multisession ES prevents threshold shifts and minimizes inflammatory reaction after acoustic overstimulation in our animal model of TTS. Consequently, auditory temporal area stimulation can be considered a potential approach to hearing preservation after mild sound overstimulation. ChAT quantitative immunocytochemistry results also indicate that TTS induces short-term neurotransmitter depletion of cholinergic terminals apposed on OHC, which recovers in the long term. Significant long-term increases in the amount of neurotransmitter in terminals (OD analysis), in the electrically stimulated experimental group, indicate persistent plastic activation of the MOC with normal thresholds, which should be explored in future research. Furthermore, the widening range of sizes and differences in OD along the cochleotopic axis suggests that chronic multisession anodal stimulation helps also to preserve the tonotopic neurotransmission of the efferent olivocochlear terminals in the inner ear.

In line with its role in cochlear amplification, activation of the efferent system induces a protective effect against noise overstimulation (Handrock and Zeisberg, 1982; Patuzzi and Thompson, 1991; Zheng et al., 1997; Tong et al., 2013; Dinh et al., 2015; Boero et al., 2018). However, in this paper, we provide data supporting new strategies based on cortical activation for preventing repetitive TTS and eventually HHL. Our multisession anodal stimulation protocol clearly avoids threshold shifts after TTS. The neurological basis for such a sustained and reversible decrease in hearing sensitivity is most likely related to the activation of long-term potentiation of Hebbian responses of the circuits involved in MOC activation (i.e., cortical, midbrain, or superior olivary circuits or most probably all of them). Exploiting neural effects of AC repetitive stimulation will enable the

development of new strategies for treating diseases with altered hearing sensitivity (hyperacusis) or hearing loss by acoustic overstimulation.

Performing repetitive chronic stimulation of the temporal cortex of patients will, nevertheless, require overcoming two obstacles: developing a non-invasive procedure and deeply stimulating the sulcus of cerebral cortex convolutions. Notwithstanding the difficulties, a new electric stimulation approach based on temporal interfering electric fields has been recently reported (Grossman et al., 2017; Sunshine et al., 2020). This procedure induces deep stimulation through surface electrodes and, therefore, is a promising method for chronic repetitive stimulation in patients, especially considering the results from ongoing experiments in our laboratory.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

ETHICS STATEMENT

The animal study was reviewed and approved by the Comite de Bioetica–University of Salamanca Servicio de Ordenación y Estructura Sanitaria Ganadera. Dirección General de Producción Agropecuaria e Infraestructuras Agrarias Consejería de Agricultura y Ganadería de la Junta de Castilla y León.

AUTHOR CONTRIBUTIONS

MM designed the experiments and wrote the manuscript. AC-R, ID, IP, VC, DP-G, and MM performed the experiments and analyzed the data. IP contributed by performing histological methods. ID conducted the quantitative immunocytochemical study and performed the statistical analysis. AC-R and DP-G contributed to the analysis of ABRs. All authors participated in the discussion of the experiments.

FUNDING

This research was supported by a grant from the University of Salamanca, Ministry of Economy and Competitiveness of the Spanish Government SAF2016-78898-C2-2-R and BFU2017-82375-R from JCYL SA070P17.

ACKNOWLEDGMENTS

This article is dedicated to the memory of Prof. Facundo Valverde. His outstanding work in neuroscience will remain forever in our memory and in that of the world of science. We would like to thank Ines Santos Fernandez del Campo for her help in counting cochlear macrophages and Prof. JM Juiz and Carlos V. Melo for their careful revision of the syntax and grammar of the manuscript.

REFERENCES

Andoh, M., Ikegaya, Y., and Koyama, R. (2019). Synaptic pruning by microglia in epilepsy. *J. Clin. Med.* 8:2170. doi: 10.3390/jcm8122170

Bajo, V. M., and Moore, D. R. (2005). Descending projections from the auditory cortex to the inferior colliculus in the gerbil, Meriones unguiculatus. *J. Comp. Neurol.* 486, 101–116. doi: 10.1002/cne.20542

Bajo, V. M., Nodal, F. R., Bizley, J. K., Moore, D. R., and King, A. J. (2007). The ferret auditory cortex: descending projections to the inferior colliculus. *Cereb. Cortex* 17, 475–491. doi: 10.1093/cercor/bhj164

Boero, L. E., Castagna, V. C., Di Guilmi, M. N., Goutman, J. D., Elgoyhen, A. B., and Gómez-Casati, M. E. (2018). Enhancement of the medial olivocochlear system prevents hidden hearing loss. *J. Neurosci.* 38, 7440–7451. doi: 10.1523/JNEUROSCI.0363-18.2018

Burda, H., Ballast, L., and Bruns, V. (1988). Cochlea in old world mice and rats (Muridae). *J. Morphol.* 198, 269–285. doi: 10.1002/jmor.1051980303

Cai, T., Seymour, M. L., Zhang, H., Pereira, F. A., and Groves, A. K. (2013). Conditional deletion of Atoh1 reveals distinct critical periods for survival and function of hair cells in the organ of Corti. *J. Neurosci.* 33, 10110–10122. doi: 10.1523/JNEUROSCI.5606-12.2013

Colmenárez-Raga, A. C., Díaz, I., Pernia, M., Pérez-González, D., Delgado-García, J. M., Carro, J., et al. (2019). Reversible functional changes evoked by anodal epidural direct current electrical stimulation of the rat auditory cortex. *Front. Neurosci.* 13:356. doi: 10.3389/fnins.2019.00356

De Paola, V., Holtmaat, A., Knott, G., Song, S., Wilbrecht, L., Caroni, P., et al. (2006). Cell type-specific structural plasticity of axonal branches and boutons in the adult neocortex. *Neuron* 49, 861–875. doi: 10.1016/j.neuron.2006.02.017

De Venecia, R. K., Liberman, M. C., Guinan, J. J., and Brown, M. C. (2005). Medial olivocochlear reflex interneurons are located in the posteroventral cochlear nucleus: a kainic acid lesion study in guinea pigs. *J. Comp. Neurol.* 487, 345–360. doi: 10.1002/cne.20550

Delano, P. H., Belkhiria, C., Vergara, R. C., Martínez, M., Leiva, A., Andrade, M., et al. (2020). Reduced suprathreshold auditory nerve responses are associated with thinner temporal cortex and slower processing speed in presbycusis. *bioRxiv* [Preprint]. doi: 10.1101/2020.02.12.945337

Dinh, C. T., Goncalves, S., Bas, E., Van De Water, T. R., and Zine, A. (2015). Molecular regulation of auditory hair cell death and approaches to protect sensory receptor cells and/or stimulate repair following acoustic trauma. *Front. Cell. Neurosci.* 9:96. doi: 10.3389/fncel.2015.00096

Doucet, J. R., Molavi, D. L., and Ryugo, D. K. (2003). The source of corticocollicular and corticobulbar projections in area Te1 of the rat. *Exp. Brain Res.* 153, 461–466. doi: 10.1007/s00221-003-1604-4

Dragicevic, C. D., Aedo, C., León, A., Bowen, M., Jara, N., Terreros, G., et al. (2015). The olivocochlear reflex strength and cochlear sensitivity are independently modulated by auditory cortex microstimulation. *J. Assoc. Res. Otolaryngol.* 16, 223–240. doi: 10.1007/s10162-015-0509-9

Elgueda, D., Delano, P. H., and Robles, L. (2011). Effects of electrical stimulation of olivocochlear fibers in cochlear potentials in the chinchilla. *J. Assoc. Res. Otolaryngol.* 12, 317–327. doi: 10.1007/s10162-011-0260-9

Eybalin, M., and Pujol, R. (1987). Choline acetyltransferase (ChAT) immunoelectron microscopy distinguishes at least three types of efferent synapses in the organ of Corti. *Exp. Brain Res.* doi: 10.1007/BF00236298

Feliciano, M., and Potashner, S. J. (2002). Evidence for a glutamatergic pathway from the guinea pig auditory cortex to the inferior colliculus. *J. Neurochem.* 65, 1348–1357. doi: 10.1046/j.1471-4159.1995.65031348.x

Fenoy, A. J., Severson, M. A., Volkov, I. O., Brugge, J. F., and Howard, M. A. (2006). Hearing suppression induced by electrical stimulation of human auditory cortex. *Brain Res.* 1118, 75–83. doi: 10.1016/j.brainres.2006.08.013

Fredelius, L., and Rask-Andersen, H. (1990). The role of macrophages in the disposal of degeneration products within the organ of corti after

acoustic overstimulation. *Acta Otolaryngol.* 109, 76–82. doi: 10.3109/00016489009107417

Frye, M. D., Ryan, A. F., and Kurabi, A. (2019). Inflammation associated with noise-induced hearing loss. *J. Acoust. Soc. Am.* 146, 4020–4032. doi: 10.1121/1.5132545

Frye, M. D., Yang, W., Zhang, C., Xiong, B., and Hu, B. H. (2017). Dynamic activation of basilar membrane macrophages in response to chronic sensory cell degeneration in aging mouse cochleae. *Hear. Res.* 344, 125–134. doi: 10.1016/j.heares.2016.11.003

Fujii, T., Mashimo, M., Moriwaki, Y., Misawa, H., Ono, S., Horiguchi, K., et al. (2017). Physiological functions of the cholinergic system in immune cells. *J. Pharmacol. Sci.* 134, 1–21. doi: 10.1016/j.jphs.2017.05.002

Galambos, R. (1956). Suppression of auditory nerve activity by stimulation of efferent fibers to cochlea. *J. Neurophysiol.* 19, 424–437. doi: 10.1152/jn.1956.19.5.424

Grillo, F. W., Song, S., Teles-Grilo Ruivo, L. M., Huang, L., Gao, G., Knott, G. W., et al. (2013). Increased axonal bouton dynamics in the aging mouse cortex. *Proc. Natl. Acad. Sci. U. S. A.* 110, E1514–E1523. doi: 10.1073/pnas.1218731110

Grossman, N., Bono, D., Dedic, N., Kodandaramaiah, S. B., Rudenko, A., Suk, H. J., et al. (2017). Noninvasive deep brain stimulation via temporally interfering electric fields. *Cell* 169, 1029–1041.e16. doi: 10.1016/j.cell.2017.05.024

Guinan, J. J. (2006). Olivocochlear efferents: anatomy, physiology, function, and the measurement of efferent effects in humans. *Ear. Hear.* 27, 589–607. doi: 10.1097/01.aud.0000240507.83072.e7

Guinan, J. J. (2010). Cochlear efferent innervation and function. *Curr. Opin. Otolaryngol. Head Neck Surg.* 18, 447–453. doi: 10.1097/MOO.0b013e32833e05d6

Handrock, M., and Zeisberg, J. (1982). The influence of the efferent system on adaptation, temporary and permanent threshold shift. *Arch. Otorhinolaryngol.* 234, 191–195. doi: 10.1007/BF00453630

He, W., Yu, J., Sun, Y., and Kong, W. (2020). Macrophages in noise-exposed cochlea: changes, regulation and the potential role. *Aging Dis.* 11, 191–199. doi: 10.14336/AD.2019.0723

Hirose, K., Discolo, C. M., Keasler, J. R., and Ransohoff, R. (2005). Mononuclear phagocytes migrate into the murine cochlea after acoustic trauma. *J. Comp. Neurol.* 489, 180–194. doi: 10.1002/cne.20619

Hong, S., Dissing-Olesen, L., and Stevens, B. (2016). New insights on the role of microglia in synaptic pruning in health and disease. *Curr. Opin. Neurobiol.* 36, 128–134. doi: 10.1016/j.conb.2015.12.004

Horváth, M., Ribári, O., Répássy, G., Tóth, I. E., Boldogkõi, Z., and Palkovits, M. (2003). Intracochlear injection of pseudorabies virus labels descending auditory and monoaminerg projections to olivocochlear cells in guinea pig. *Eur. J. Neurosci.* 18, 1439–1447. doi: 10.1046/j.1460-9568.2003.02870.x

Kawaguchi, Y. (2017). Pyramidal cell subtypes and their synaptic connections in layer 5 of rat frontal cortex. *Cereb. Cortex* 27, 5755–5771. doi: 10.1093/cercor/bhx252

Kim, M. S., Koo, H., Han, S. W., Paulus, W., Nitsche, M. A., Kim, Y. H., et al. (2017). Repeated anodal transcranial direct current stimulation induces neural plasticity-associated gene expression in the rat cortex and hippocampus. *Restor. Neurol. Neurosci.* 35, 137–146. doi: 10.3233/RNN-160689

Kokko-Cunningham, A., and Ades, H. W. (1976). Acetylcholinesterase activity in the chinchillaorgan of corti in normal and acoustically overstimulated animals. *Acta Otolaryngol.* 81, 48–56. doi: 10.3109/00016487609107476

Kotak, V. C., Breithaupt, A. D., and Sanes, D. H. (2007). Developmental hearing loss eliminates long-term potentiation in the auditory cortex. *Proc. Natl. Acad. Sci. U. S. A.* 104, 3550–3555. doi: 10.1073/pnas.0607177104

Kuhar, M. J. (1979). Sodium-dependent high affinity choline uptake. *Prog. Brain Res.* 30, 15–21. doi: 10.1016/S0079-6123(08)64622-1

Kujawa, S. G., and Liberman, M. C. (2009). Adding insult to injury: cochlear nerve degeneration after "temporary" noise-induced hearing loss. *J. Neurosci.* 29, 14077–14085. doi: 10.1523/JNEUROSCI.2845-09.2009

Lamas, V., Alvarado, J. C., Carro, J., and Merchán, M. A. (2013). Long-term evolution of brainstem electrical evoked responses to sound after restricted ablation of the auditory cortex. *PLoS One* 8:e73585. doi: 10.1371/journal.pone.0073585

Lamas, V., Arevalo, J. C., Juíz, J. M., and Merchán, M. A. (2014). Acoustic input and efferent activity regulate the expression of molecules involved in cochlear micromechanics. *Front. Syst. Neurosci.* 8:253. doi: 10.3389/fnsys.2014.00253

Lamas, V., Esteévez, S., Perniía, M., Plaza, I., and Merchaín, M. A. (2017). Stereotactically-guided ablation of the rat auditory cortex, and localization of the lesion in the brain. *J. Vis. Exp.* 128:56429. doi: 10.3791/56429

León, A., Elgueda, D., Silva, M. A., Hamamé, C. M., and Delano, P. H. (2012). Auditory cortex basal activity modulates cochlear responses in chinchillas. *PLoS One* 7:e36203. doi: 10.1371/journal.pone.0036203

Liberman, M. C., Epstein, M. J., Cleveland, S. S., Wang, H., and Maison, S. F. (2016). Toward a differential diagnosis of hidden hearing loss in humans. *PLoS One* 11:e0162726. doi: 10.1371/journal.pone.0162726

Lobarinas, E., Spankovich, C., and Le Prell, C. G. (2017). Evidence of "hidden hearing loss" following noise exposures that produce robust TTS and ABR wave-I amplitude reductions. *Hear. Res.* 349, 155–163. doi: 10.1016/j.heares.2016.12.009

Lopez-Poveda, E. A. (2018). Olivocochlear efferents in animals and humans: from anatomy to clinical relevance. *Front. Neurol.* 9:197. doi: 10.3389/fneur.2018.00197

Lund, R. D., and Lund, J. S. (1976). Plasticity in the developing visual system: the effects of retinal lesions made in young rats. *J. Comp. Neurol.* 169, 133–154. doi: 10.1002/cne.901690202

Malmierca, M. S., and Merchán, M. A. (2004). "Auditory System," in The Rat Nervous System, ed. G. Paxinos (San Diego, CA: Academic Press), 995–1080.

Malmierca, M. S., Le Beau, F. E., and Rees, A. (1996). The topographical organization of descending projections from the central nucleus of the inferior colliculus in guinea pig. *Hear. Res.* 93, 167–180. doi: 10.1016/0378-5955(95)00227-8

Malmierca, M. S., and Ryugo, D. K. (2011). "Descending connections of auditory cortex to the midbrain and brain stem," in *The Auditory Cortex*, eds J. Winer, and C. Schreiner (Boston, MA: Springer), 189–208. doi: 10.1007/978-1-4419-0074-6

Matsuo, A., Bellier, J. P., Nishimura, M., Yasuhara, O., Saito, N., and Kimura, H. (2011). Nuclear choline acetyltransferase activates transcription of a high-affinity choline transporter. *J. Biol. Chem.* 286, 5836–5845. doi: 10.1074/jbc.M110.147611

Mellott, J. G., Bickford, M. E., and Schofield, B. R. (2014). Descending projections from auditory cortex to excitatory and inhibitory cells in the nucleus of the brachium of the inferior colliculus. *Front. Syst. Neurosci.* 8:188. doi: 10.3389/fnsys.2014.00188

Mounier-Kuhn, P., and Haguenauer, J. P. (1967). Variations de L'activité cholinestérasique dans les cellules ciliées de L'organe de corti après exposition á des sons de fréquence basse ou elevée. *Acta Otolaryngol.* 63, 297–303. doi: 10.3109/00016486709128763

Nitsche, M. A., Cohen, L. G., Wassermann, E. M., Priori, A., Lang, N., Antal, A., et al. (2008). Transcranial direct current stimulation: state of the art 2008. *Brain Stimul.* 1, 206–223. doi: 10.1016/j.brs.2008.06.004

Nitsche, M. A., and Paulus, W. (2000). Excitability changes induced in the human motor cortex by weak transcranial direct current stimulation. *J. Physiol.* 527 Pt 3, 633–639. doi: 10.1109/ICC.1999.765531

Pasantes-Morales, H., and Tuz, K. (2006). Volume changes in neurons: hyperexcitability and neuronal death. *Contrib. Nephrol.* 152, 221–240. doi: 10.1159/000096326

Pasaoglu, T., and Schikorski, T. (2016). Presynaptic size of associational/commissural CA3 synapses is controlled by fibroblast growth factor 22 in adult mice. *Hippocampus* 26, 151–160. doi: 10.1002/hipo.22499

Patuzzi, R. B., and Thompson, M. L. (1991). Cochlear efferent neurones and protection against acoustic trauma: protection of outer hair cell receptor current and interanimal variability. *Hear. Res.* 54, 45–58. doi: 10.1016/0378-5955(91)90135-V

Paxinos, G., and Watson, C. (2005). *The Rat Brain in Stereotaxic Coordinates*. 5th edn. Cambridge, MA: Elsevier Academic Press.

Pernia, M., Díaz, I., Colmenárez-Raga, A. C., Rivadulla, C., Cudeiro, J., Plaza, I., et al. (2020). Cross-modal reaction of auditory and visual cortices after long-term bilateral hearing deprivation in the rat. *Brain Struct. Funct.* 225, 129–148. doi: 10.1007/s00429-019-01991-w

Perrot, X., Ryvlin, P., Isnard, J., Guénot, M., Catenoix, H., Fischer, C., et al. (2006). Evidence for corticofugal modulation of peripheral auditory activity in humans. *Cereb. Cortex* 16, 941–948. doi: 10.1093/cercor/bhj035

Petrof, I., and Sherman, S. M. (2013). Functional significance of synaptic terminal size in glutamatergic sensory pathways in thalamus and cortex. *J. Physiol.* 591, 3125–3131. doi: 10.1113/jphysiol.2012.247619

Pierce, J. P., and Milner, T. A. (2001). Parallel increases in the synaptic and surface areas of mossy fiber terminals following seizure induction. *Synapse* 39, 249–256. doi: 10.1002/1098-2396(20010301)39:3<249::AID-SYN1006<3.0.CO;2-5

Saldaña, E., Feliciano, M., and Mugnaini, E. (1996). Distribution of descending projections from primary auditory neocortex to inferior colliculus mimics the topography of intracollicular projections. *J. Comp. Neurol.* 371, 15–40. doi: 10.1002/(sici)1096-9861(19960715)371:1<15::aid-cne2>3.0.co;2-o

Sammons, R. P., Clopath, C., and Barnes, S. J. (2018). Size-dependent axonal bouton dynamics following visual deprivation in vivo. *Cell Rep.* 22, 576–584. doi: 10.1016/j.celrep.2017.12.065

Schofield, B. R., and Cant, N. B. (1999). Descending auditory pathways: projections from the inferior colliculus contact superior olivary cells that project bilaterally to the cochlear nuclei. *J. Comp. Neurol.* 409, 210–223. doi: 10.1002/(sici)1096-9861(19990628)409:2<210::aid-cne3>3.0.co;2-a

Sellgren, C. M., Gracias, J., Watmuff, B., Biag, J. D., Thanos, J. M., Whittredge, P. B., et al. (2019). Increased synapse elimination by microglia in schizophrenia patient-derived models of synaptic pruning. *Nat. Neurosci.* 22, 374–385. doi: 10.1038/s41593-018-0334-7

Senatorov, V. V., and Hu, B. (2002). Extracortical descending projections to the rat inferior colliculus. *Neuroscience* 115, 243–250. doi: 10.1016/s0306-4522(02)00316-0

Simon, J. R., and Kuhar, M. J. (1976). High affinity choline uptake: ionic and energy requirements. *J. Neurochem.* 27, 93–99. doi: 10.1111/j.1471-4159.1976.tb01549.x

Spangler, K. M., Cant, N. B., Henkel, C. K., Farley, G. R., and Warr, W. B. (1987). Descending projections from the superior olivary complex to the cochlear nucleus of the cat. *J. Comp. Neurol.* 259, 452–465. doi: 10.1002/cne.902590311

Stanic, D., Parish, C. L., Zhu, W. M., Krstew, E. V., Lawrence, A. J., Drago, J., et al. (2003). Changes in function and ultrastructure of striatal dopaminergic terminals that regenerate following partial lesions of the SNpc. *J. Neurochem.* 86, 329–343. doi: 10.1046/j.1471-4159.2003.01843.x

Statman, A., Kaufman, M., Minerbi, A., Ziv, N. E., and Brenner, N. (2014). Synaptic size dynamics as an effectively stochastic process. *PLoS Comput. Biol.* 10:e1003846. doi: 10.1371/journal.pcbi.1003846

Straka, M. M., Hughes, R., Lee, P., and Lim, H. H. (2015). Descending and tonotopic projection patterns from the auditory cortex to the inferior colliculus. *Neuroscience* 300, 325–337. doi: 10.1016/j.neuroscience.2015.05.032

Sunshine, M. D., Cassarà, A. M., Neufeld, E. M., Grossman, N., Otto, K. J., Boyden, E. S., et al. (2020). Phrenic motor neuron activation using temporal interference stimulation. *FASEB J.* 34, 1–1. doi: 10.1096/fasebj.2020.34.s1.06546

Terreros, G., and Delano, P. H. (2015). Corticofugal modulation of peripheral auditory responses. *Front. Syst. Neurosci.* 9:134. doi: 10.3389/fnsys.2015.00134

Thompson, A. M., and Schofield, B. R. (2000). Afferent projections of the superior olivary complex. *Microsc. Res. Tech.* 51, 330–354. doi: 10.1002/1097-0029(20001115)51:4<330::AID-JEMT4<3.0.CO;2-X

Tong, H., Kopp-Scheinpflug, C., Pilati, N., Robinson, S. W., Sinclair, J. L., Steinert, J. R., et al. (2013). Protection from noise-induced hearing loss by Kv2.2 potassium currents in the central medial olivocochlear system. *J. Neurosci.* 33, 9113–9121. doi: 10.1523/JNEUROSCI.5043-12.2013

Vetter, D. E., Saldaña, E., and Mugnaini, E. (1993). Input from the inferior colliculus to medial olivocochlear neurons in the rat: a double label study with PHA-L and cholera toxin. *Hear. Res.* 70, 173–186. doi: 10.1016/0378-5955(93)90156-u

Warr, W. B., and Boche, J. E. (2003). Diversity of axonal ramifications belonging to single lateral and medial olivocochlear neurons. *Exp. Brain Res.* 153, 499–513. doi: 10.1007/s00221-003-1682-3

Weedman, D. L., and Ryugo, D. K. (1996). Projections from auditory cortex to the cochlear nucleus in rats: synapses on granule cell dendrites. *J. Comp. Neurol.* 371, 311–324. doi: 10.1002/(sici)1096-9861(19960722)371:2<311::aid-cne10>3.0.co;2-v

Winer, J. A., Larue, D. T., Diehl, J. J., and Hefti, B. J. (1998). Auditory cortical projections to the cat inferior colliculus. *J. Comp. Neurol.* 400, 147–174. doi: 10.1002/(sici)1096-9861(19981019)400:2<147::aid-cne1>3.0.co;2-9

Wood, M. B., and Zuo, J. (2017). The contribution of immune infiltrates to ototoxicity and cochlear hair cell loss. *Front. Cell. Neurosci.* 11:106. doi: 10.3389/fncel.2017.00106

Xiao, Z., and Suga, N. (2002). Modulation of cochlear hair cells by the auditory cortex in the mustached bat. *Nat. Neurosci.* 5, 57–63. doi: 10.1038/nn786

Zhang, J. (2013). Auditory cortex stimulation to suppress tinnitus: mechanisms and strategies. *Hear. Res.* 295, 38–57. doi: 10.1016/j.heares.2012.05.007

Zheng, X. Y., Henderson, D., McFadden, S. L., and Hu, B. H. (1997). The role of the cochlear efferent system in acquired resistance to noise-induced hearing loss. *Hear. Res.* 104, 191–203. doi: 10.1016/S0378-5955(96)00187-6

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Díaz, Colmenarez-Raga, Pérez-González, Carmona, Plaza Lopez and Merchán. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Neuroscience

REVIEW
published: 03 August 2021
doi: 10.3389/fnins.2021.690223

When and How Does the Auditory Cortex Influence Subcortical Auditory Structures? New Insights About the Roles of Descending Cortical Projections

*Samira Souffi[1], Fernando R. Nodal[2], Victoria M. Bajo[2] and Jean-Marc Edeline[1]**

[1] Department of Integrative and Computational Neurosciences, Paris-Saclay Institute of Neuroscience (NeuroPSI), UMR CNRS 9197, Paris-Saclay University, Orsay, France, [2] Department of Physiology, Anatomy and Genetics, Medical Sciences Division, University of Oxford, Oxford, United Kingdom

For decades, the corticofugal descending projections have been anatomically well described but their functional role remains a puzzling question. In this review, we will first describe the contributions of neuronal networks in representing communication sounds in various types of degraded acoustic conditions from the cochlear nucleus to the primary and secondary auditory cortex. In such situations, the discrimination abilities of collicular and thalamic neurons are clearly better than those of cortical neurons although the latter remain very little affected by degraded acoustic conditions. Second, we will report the functional effects resulting from activating or inactivating corticofugal projections on functional properties of subcortical neurons. In general, modest effects have been observed in anesthetized and in awake, passively listening, animals. In contrast, in behavioral tasks including challenging conditions, behavioral performance was severely reduced by removing or transiently silencing the corticofugal descending projections. This suggests that the discriminative abilities of subcortical neurons may be sufficient in many acoustic situations. It is only in particularly challenging situations, either due to the task difficulties and/or to the degraded acoustic conditions that the corticofugal descending connections bring additional abilities. Here, we propose that it is both the top-down influences from the prefrontal cortex, and those from the neuromodulatory systems, which allow the cortical descending projections to impact behavioral performance in reshaping the functional circuitry of subcortical structures. We aim at proposing potential scenarios to explain how, and under which circumstances, these projections impact on subcortical processing and on behavioral responses.

Keywords: auditory processing, corticofugal projections, inferior colliculus, degraded acoustic conditions, neuromodulation, frontal cortex, auditory plasticity, active listening

OPEN ACCESS

Edited by:
Erika Skoe,
University of Connecticut, United States

Reviewed by:
Josef Syka,
Institute of Experimental Medicine (ASCR), Czechia
Paul Hinckley Delano,
University of Chile, Chile

****Correspondence:***
Jean-Marc Edeline
jean-marc.edeline@u-psud.fr

Specialty section:
This article was submitted to Auditory Cognitive Neuroscience, a section of the journal Frontiers in Neuroscience

Received: *02 April 2021*
Accepted: *14 June 2021*
Published: *03 August 2021*

Citation:
Souffi S, Nodal FR, Bajo VM and Edeline J-M (2021) When and How Does the Auditory Cortex Influence Subcortical Auditory Structures? New Insights About the Roles of Descending Cortical Projections. Front. Neurosci. 15:690223. doi: 10.3389/fnins.2021.690223

INTRODUCTION

The auditory cortex has been viewed as the ultimate step in processing the rich acoustic stream constantly reaching our ears and also as a key structure in cognitive tasks involving auditory stimuli (Weinberger and Diamond, 1987; Edeline, 1999; Weinberger, 2004; Ohl and Scheich, 2005; Fritz et al., 2007). Indeed, the plasticity of auditory cortex network has been described in many situations

ranging from frequency discrimination (Edeline and Weinberger, 1993; Edeline et al., 1993; Fritz et al., 2003, 2005) or spatial discrimination tasks (Lee and Middlebrooks, 2011; Wood et al., 2019) to pitch extraction (Bizley et al., 2013), attentional tasks (Otazu et al., 2009), selective attention (Wittekindt et al., 2014), and predictive coding (Malmierca et al., 2015).

Besides its role in cognitive functions, several recent studies performed on different species have promoted the idea that auditory cortex is also a key structure in building noise-invariant representations of communication sounds (Narayan et al., 2007; Carruthers et al., 2013, 2015; Rabinowitz et al., 2013; Schneider and Woolley, 2013; Mesgarani et al., 2014; Ni et al., 2017; Aushana et al., 2018; Beetz et al., 2018; Town et al., 2018; Souffi et al., 2020). For example, the cortical responses to conspecific vocalizations, and their discriminations by cortical neurons were largely preserved during various types of acoustic alterations performed in the spectral and temporal domain (Souffi et al., 2020).

In this review, we propose new roles of descending cortical projections reaching the auditory thalamus and the inferior colliculus. These two subcortical structures receive the dominant part of the corticofugal inputs and had been explored in a large number of species and under different listening conditions. Therefore, we will focus on the specific effects mediated by those circuits without forgetting that the effects of the cortical descending projections can also modify earlier relay stations. We will describe studies from different animal species (mice, rats, guinea pigs, ferrets, bats, and birds). While the descending cortical projections in the auditory system are potentially equivalent in all species, the frontal circuitry could strongly vary between species making generalization of results more difficult.

In the present review, we will first describe the extent to which cortical neurons robustly code the representation of target stimuli in acoustically challenging conditions. Next, we will examine data suggesting that noise-invariant representations do also exist in subcortical auditory structures. In the last sections of the review, we will point out that, despite numerous experiments which aimed at describing the influence of corticofugal connections at the thalamic and collicular level, it is only the use of cell-targeted activation/inactivation methodologies combined with behavioral tasks that have recently unraveled whether the auditory cortex impacts on subcortical processing in challenging conditions.

EVIDENCE FOR NOISE-INVARIANT REPRESENTATIONS IN AUDITORY CORTEX

Our ears are constantly bombarded by a complex sound mixture, which generates challenging acoustic conditions for speech understanding. These degraded acoustic conditions can be the presence of reverberations, for example created by the shape, size, and objects in the room in closed spaces, the presence of concomitant sound sources with the particular case of the "cocktail party" noise where a target source has to be segregated from other competing sounds (e.g., see Narayan et al., 2007) but also particular environmental conditions that can attenuate specific frequencies from the signal spectra (Mesgarani et al., 2014; Fuglsang et al., 2017; Bidelman et al., 2018). All these factors lead to difficulties in perceiving target sounds such as speech, communication sounds and music in normal-hearing subjects, but cause even more difficulties for subjects with mild to moderate hearing loss, and are very penalizing for subjects with cochlear implants, a neuroprosthetic device which restores hearing in people suffering from profound deafness. Note also that for patients with cochlear implants, the descending cortical projections to the thalamus and to the inferior colliculus are preserved but the indirect cortical modulation to the auditory periphery is lacking.

Understanding what are the spectro-temporal acoustic cues used by human subjects necessary for auditory perception in challenging conditions and the neuronal mechanisms allowing the auditory system to extract relevant cues for discriminating sounds in those acoustic conditions are major aims in psychoacoustic and auditory neuroscience.

Over the last two decades, most of the studies describing the physiological consequences of adding noise on the neuronal responses to target stimuli have been performed at the level of the primary auditory cortex (A1). In their initial study, Nagarajan et al. (2002) reported that white noise addition reduced auditory responses to conspecific communication sounds (marmoset calls) only at a 0 dB signal to noise ratio (SNR), the lowest SNR tested. This study also pointed out that cortical neurons are particularly robust to spectral degradations since there was little change in evoked responses at presentation of vocoded vocalizations [an artificial signal-processing distortion that remove the spectral content and the frequency modulation (FM) cues but partially preserved the amplitude modulation (AM) cues], even in response to only 2-band vocoded vocalizations. In contrast, temporal-envelope degradations strongly reduced the evoked firing rate and the neural synchronization to the vocalization envelope. Importantly, bandpass filtering the vocalizations between 2-30 Hz did not reduce the firing rate and neural synchronization to the vocalization envelope. Similarly, subsequent studies did not find much alterations of cortical responses for speech-like sounds presented in noise: for example, Shetake et al. (2011) in rats did not find significant reduction in neural discrimination using an index of neuronal population performance at a +12 dB SNR; the neural performance fell close to the chance level only at −12 dB SNR, the lowest SNR tested. In the field L in birds (homologous to primary auditory cortex), the neural discrimination performance was maintained down to a +5 dB SNR (Narayan et al., 2007).

Recent studies in guinea pigs have confirmed that the responses of auditory cortex neurons are particularly resistant to spectral degradations of communication sounds (such as vocoded vocalizations, e.g., Souffi et al., 2020), even in the presence of masking noise (Aushana et al., 2018). At the level of small cortical populations (2–16 simultaneous recordings), the ability to discriminate between conspecific vocalizations remained almost intact despite strong spectral alterations (Aushana et al., 2018; Souffi et al., 2020).

However, analyzing in more detail the responses of individual recordings across several signal-to-noise ratios revealed strikingly

different categories (Ni et al., 2017: marmoset; Souffi et al., 2021: guinea pigs), which ranged from neuronal responses robust to noise and specific to target stimuli, to neuronal responses sensitive to noise and specific to masking noises. In fact, the initial results of Bar-Yosef and Nelken (2007) in the cat primary auditory cortex have already pointed out that some cortical neurons can be more specific to the background noise than to the actual communication sounds. In addition, context seems to be important too, and neurons assigned to a particular category can change category depending on the type of noise, indicating that different types of masking noise activate different subpopulations of neurons in the auditory cortex and subcortical auditory structures (Ni et al., 2017; Souffi et al., 2021).

Several hypotheses have been formulated to account for the performance of auditory cortex neurons in detecting target stimuli in masking noise. For example, it was proposed that noise tolerance is correlated with adaptation to the stimulus statistics, which is more pronounced at the cortical than at the subcortical level in ferrets (Rabinowitz et al., 2013). A dynamic model of synaptic depression was also suggested as a potential mechanism for robust speech representation in the human auditory cortex (Mesgarani et al., 2014). Alternatively, a simple feedforward inhibition circuit operating in a sparse coding scheme was viewed as a mechanism to explain background-invariant responses detected for a population of neurons in the zebra finch secondary auditory cortex (Schneider and Woolley, 2013).

As we will see below, it is important to determine whether these mechanisms only operate at cortical level or whether they are general mechanisms operating at all the levels of the central auditory system.

SUBCORTICAL IMPLICATIONS IN BUILDING NOISE-INVARIANT REPRESENTATIONS

Compared with the large literature focused on the auditory cortex, only a few studies have described the resistance to noise of subcortical neurons. Nonetheless, a direct comparison between the consequences of acoustic degradation in different structures is the most straightforward way for dissecting where invariant representations emerged. At the thalamic level, a massive reduction in firing rate and temporal reliability of evoked responses was reported in rats during the noise condition when target stimuli and background noise were at the same intensity level (0 dB SNR, Martin et al., 2004). In the avian auditory system, Schneider and Woolley (2013) described the emergence of noise-invariant responses for a subset of cells (the broad spike cells) of a secondary auditory area (area NCM), whereas neurons in the field L and the mesencephalicus lateralis dorsalis (homologous of the primary auditory cortex and inferior colliculus, respectively) show background-corrupted responses. They proposed that a sparse coding scheme (in the sense that neurons show less driven response to the same stimulus and respond only to a small subset of the stimuli) operating within the area NCM allows the emergence of this noise-invariant representation. Note that, in rats, such a sparse representation already exists as early as A1 (Hromádka et al., 2008).

Noise-invariant representations were also reported in A1 of anesthetized ferrets (Rabinowitz et al., 2013). This study suggested a progressive emergence of noise-invariant responses from the auditory nerve to the inferior colliculus (IC) and to A1, and proposed the adaptation to the noise statistics as a key mechanism to account for the noise-invariant representation in A1. However, Lohse et al. (2020) in mice have recently challenged this view. Indeed, they showed that collicular, thalamic and cortical neurons display similar contrast gain control with the slowest time constants in A1 and importantly, the silencing of auditory cortex, did not affect the contrast gain control capacity of neurons in the inferior colliculus or in the medial geniculate body (MGB). Previous studies have already shown adaptation to stimulus intensity of subcortical neurons. First, adaptations of IC neurons to the average stimulus intensity, stimulus variance and bimodality have already been described in guinea pigs with a temporal decay of about 160 ms at 75 dB sound pressure level (SPL, Dean et al., 2005, 2008). Second, adaptation to the noise statistics shifted the temporal modulation function (TMF) of IC neurons to slower modulations, sometimes transforming band-pass TMF to low pass TMF in about 200 ms of noise presentation (gerbils: Lesica and Grothe, 2008).

In fact, Nelken et al. (1999) in cats have previously shown that the addition of low intensity sounds interrupts the phase locking of A1 neurons to the envelope of slowly fluctuating noise (about 10 Hz). This phenomenon has been called "locking suppression." Moreover, the high sensitivity of this suppression, occurring at intensities lower than the neuron's threshold (at −15 or −35 dB SNR), seems to be a marked phenomenon at the cortical level, present for only about half of the neurons of the MGB and absent at the level of the IC. The conclusion is that, although the detection of pure tones in fluctuating noise is possible from the IC, the segregation between the representation of sound as a perceptual object separate from noise is more explicit/complete at the cortical level. It should be noted that intracellular recordings did not reveal a particular role of cortical inhibition in the phenomenon of "locking suppression," it is already detected in the excitatory inputs received by cortical neurons (cats: Las et al., 2005; rats: Hershenhoren and Nelken, 2017).

From recordings obtained in anesthetized guinea pigs in the cochlear nucleus, inferior colliculus, auditory thalamus, A1 and a non-primary auditory cortex, Souffi et al. (2020) reported that higher discrimination performance and more accurate representations in degraded acoustic conditions (presence of masking noise or vocoding) were found in IC and MGB; cortical representations, although less accurate as the subcortical ones, were barely affected under these degraded conditions (**Figure 1**, modified from Souffi et al., 2020). Furthermore, when neuronal responses in noise were classified among a continuum in five categories from the most robust to noise (signal-like responses) to the most sensitive to noise (masker-like responses, representing accurately the masking noise), it was found, in two noise types, that these categories were distributed in the whole auditory system, with higher proportions of robust responses in inferior colliculus and thalamus (**Figure 2**, modified from

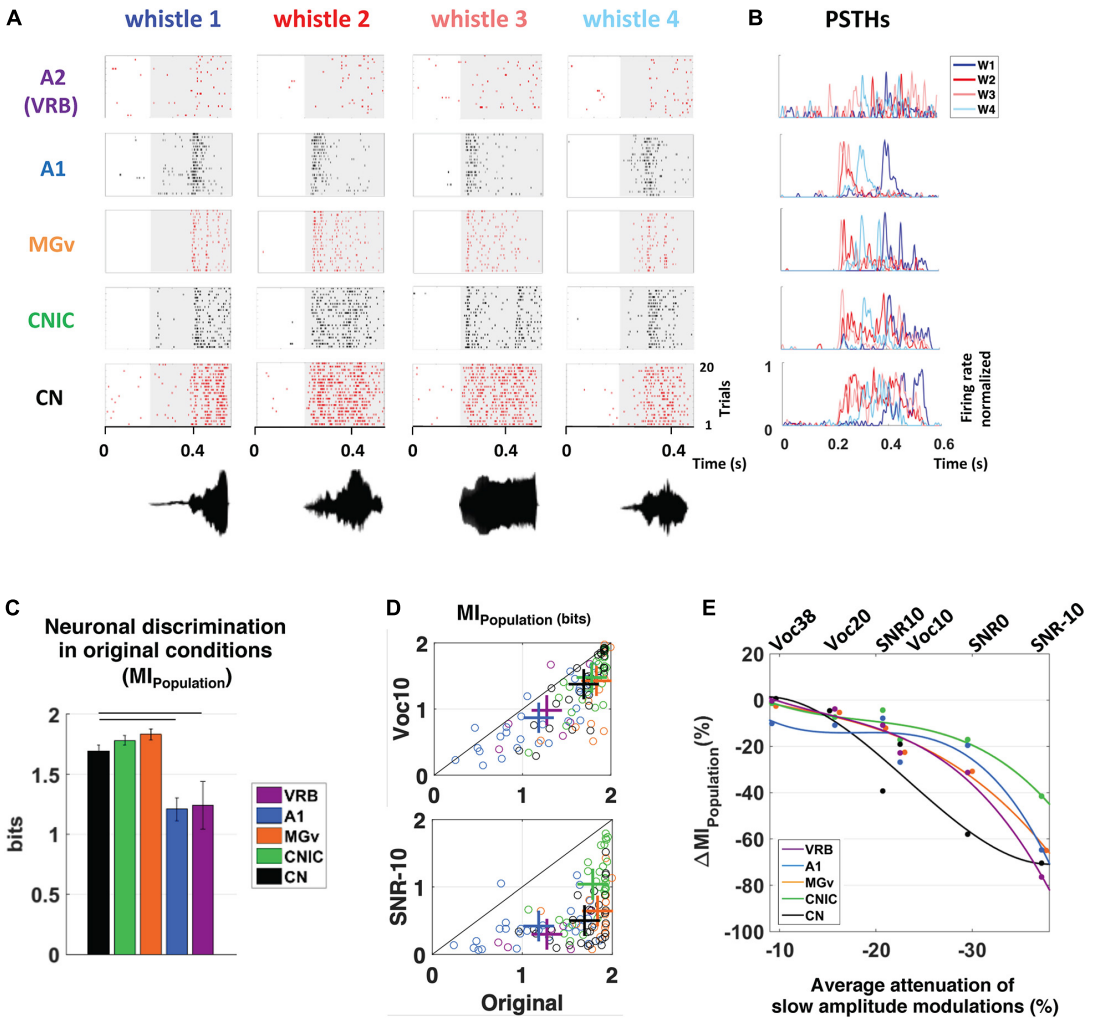

FIGURE 1 | Subcortical neurons better discriminate the vocalizations in quiet as well as in degraded conditions and alterations of slow amplitude modulations are crucial cues for explaining the decrease in discrimination performance at the subcortical and cortical levels. **(A)** From bottom to top, raster plots presenting the neuronal responses recorded in CN, CNIC, MGv, A1, and VRB. Each dot represents an action potential and each line the presentation of one of four original whistles. The gray areas correspond to the evoked activity. The waveforms of the four original whistles are displayed under the raster plots. **(B)** Peristimulus time histograms (PSTHs) of the neuronal responses presented in **(A)**. For all neuronal recordings, the four PSTHs corresponding to the four original whistles have been overlayed. **(C)** The mean values of the neuronal discrimination at the population level ($MI_{Population}$, bits) are presented for populations of 9 simultaneous multiunit recordings obtained with the four original vocalizations in CN (in black), CNIC (in green), MGv (in orange), A1 (in blue), and VRB (in purple). Error bars represent the SE of the mean and horizontal black lines represent the statistically significant differences. Note that, all the subcortical structures discriminate better the original vocalizations than cortical areas. **(D)** Scattergrams showing the modest decrease in $MI_{Population}$ (bits) with the most severe vocoded condition (Voc10, top panel) compared to the strong decrease with the most severe noisy condition (SNR-10, bottom panel). Each cross represents the mean $MI_{Population}$ obtained in degraded and original conditions. **(E)** Percentage of alterations in neuronal population discrimination abilities ($\Delta MI_{Population}$) as a function of the alterations in slow amplitude modulations induced by vocoding (Voc38, Voc20, and Voc10) or by the addition of stationary noise (SNR10, SNR0, and SNR-10). Each dot represents neuronal data ($\Delta MI_{Population}$) in CN (in black), CNIC (in green), MGv (in orange), A1 (in blue) and VRB (in purple). Polynomial curves fitting all acoustic conditions have been generated (color lines). In all conditions (vocoding or noise), there is a limit of AM reduction from which the $\Delta MI_{Population}$ decreases in cortical and subcortical structures. Thus, the reduction of slow AM cues is one of the factors explaining the neuronal discrimination performance at the subcortical and cortical levels. Modified from Souffi et al. (2020). CN, cochlear nucleus; CNIC, central nucleus of the inferior colliculus; MGv, ventral division of the medial geniculate nucleus; VRB, ventrorostral belt (secondary auditory cortex).

Souffi et al., 2021). In addition, the responses to the signal alone and to the noise alone allowed the assignment of a given recording to one of five categories to be predicted up to 70%. A link between inferior colliculus activity and behavior was pointed out in two studies showing that a tone-versus-noise discrimination task modulates the neuronal activity as early as the inferior colliculus (Slee and David, 2015; Shaheen et al., 2020). In the first one in ferrets, it was found that in the active condition, collicular responses to reference sounds were mostly suppressed and this effect was frequency-dependent with lower suppression when the target frequency was away to the Best Frequency (BF) of the neuron than when was closer. The second study

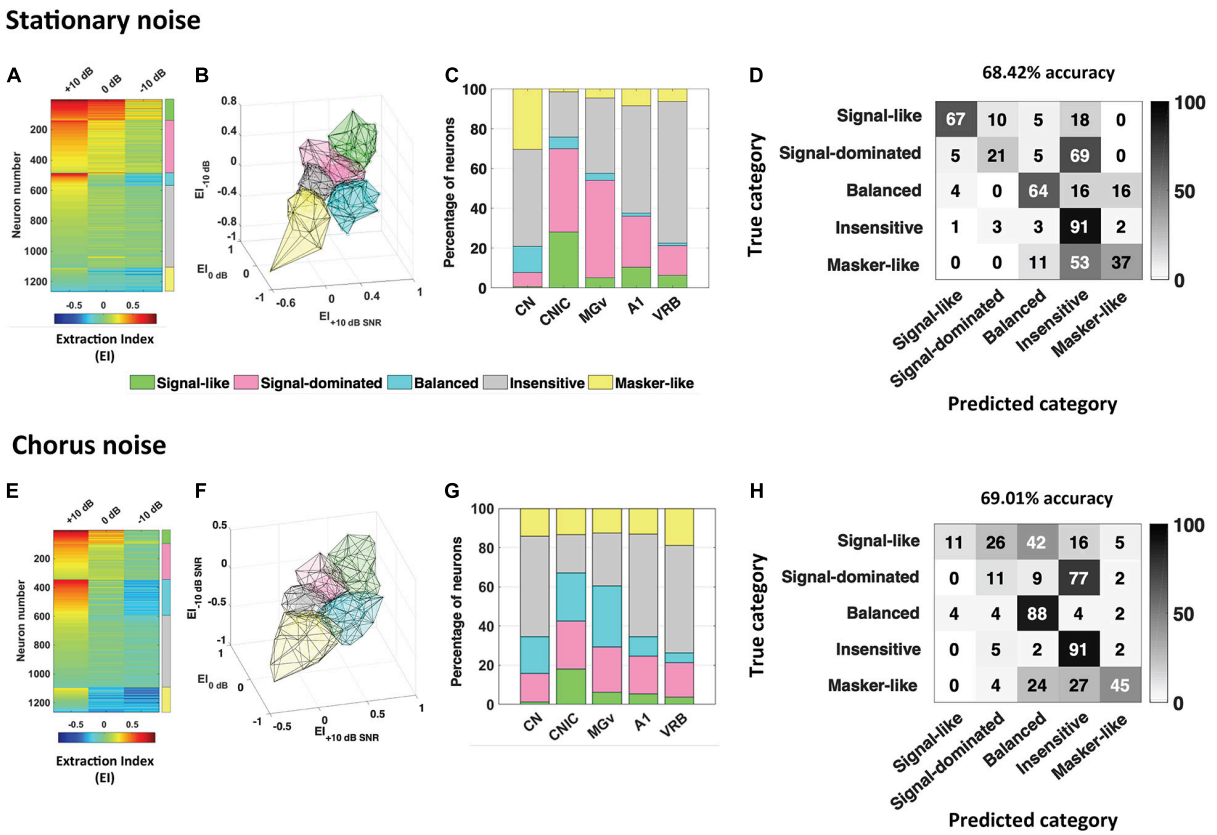

FIGURE 2 | Robustness to noise is a distributed and predictable property in the whole auditory system. The extraction index (EI) quantifies to what extent the evoked response at a given SNR is similar the response to vocalizations in quiet or to noise alone. **(A)** Each row corresponds to the extraction Index (EI) profile of a given neuronal recording obtained in the five auditory structures (CN, CNIC, MGv, A1, and VRB) in stationary noise with a color code from blue to red when progressing from low to high EI values. The color coded column on the right, delineate the identity for the five categories of responses found, "signal-like" in green, "signal-dominated" in pink, "balanced" in turquoise, "insensitive" in gray, and "masker-like" in yellow. **(B,C)** 3D representation of the five categories in stationary noise **(B)**, and proportion of each category in the five auditory structures from CN to VRB **(C)**. **(D)** Confusion matrix obtained with descriptors extracted from pure tone responses, signal alone and stationary noise alone responses. Each row corresponds to a true category and each column corresponds to a predicted category. The numbers in the confusion matrix correspond to the percentage of recordings of a given true category which have been predicted to belong to a given predicted category. Around 70% of accuracy (68.42%) was reached with these descriptors. **(E–H)** Same representations as in **(A–D)** for the responses collected in the chorus noise. Modified from Souffi et al. (2021).

quantified the neuronal discriminability in a tone-masking noise task (0 dB SNR) of IC neurons in marmoset (non-lemniscal IC: dorsal and external cortices, and lemniscal IC: central nucleus) and indicated that non-lemniscal IC neurons enhanced their neuronal discriminability in active condition whereas lemniscal IC neurons did not.

All the results together suggest that noise-invariant representations emerge very early in the auditory system under conditions of anesthetized or awake passive listening, without necessarily the involvement of cortical activity (Lohse et al., 2020).

EFFECTS OF THE CORTICOFUGAL DESCENDING PROJECTIONS

A myriad of anatomical studies have described in great detail the corticofugal projections originating from auditory cortex reaching the different subcortical relays (for reviews see Winer, 2006; Winer and Lee, 2007; Malmierca and Ryugo, 2011), but only a limited set of studies have reported the physiological effects of these projections. In this review, we focus on the descending cortical projections to the thalamus and the inferior colliculus but it should be kept in mind that descending cortical projections have been anatomically described in the dorsal cochlear nucleus (Jacomme et al., 2003). Also, the activity of the auditory nerve and the cochlea could be modulated via the olivocochlear neurons (Aedo et al., 2016; for reviews: Terreros and Delano, 2015; Elgueda and Delano, 2020) that receive direct projections from the auditory cortex (rats: Mulders and Robertson, 2000; Doucet et al., 2002; guinea pig: Coomes and Schofield, 2004; Brown et al., 2013) and the inferior colliculus (Thompson and Thompson, 1993).

For the purpose of the present review, it is particularly important to distinguish the conditions during which these effects have been reported. Some of these studies performed in

anesthetized animals have either activated or inactivated auditory cortex neurons and looked for the physiological consequences on the neuronal responses collected in subcortical auditory structures. Other studies, performed in awake behaving animals, have looked at the consequence of silencing the auditory cortex on the animal behavioral performance.

Auditory Cortical Manipulations in Anesthetized Animals

The rational of the electrophysiological experiments performed in anesthetized animals was simply to record in subcortical structures during either inactivation or electrical activation of the auditory cortex. The initial topography of the corticocollicular pathway has been described in cats by Anderson et al. (1980) combining recordings in the primary auditory cortex (A1), the anterior auditory field (AAF) and the secondary auditory cortex (AII) with anterograde (3H*Leucine) tracer injections and showing labeled terminals in IC, including the central nucleus where the changes in position of the labeling agreed with the tonotopic axes of the central nucleus of the IC (CNIC) and the tuning frequency of the neurons recorded at the injection sites. The glutamatergic nature of this pathway was suggested by Feliciano and Potashner (1995) after ablation of the auditory cortex in guinea pigs and determination of the uptake and release of radioactive Aspartate in the inferior colliculus. Initial experiments in cats have silenced the entire auditory cortex by cooling and have reported both excitatory and inhibitory effects on responses of auditory thalamus neurons (Ryugo and Weinberger, 1976) and in the inferior colliculus. In many cases, "On" responses were unaffected whereas long latencies responses were largely reduced (see also in rats, Cotillon and Edeline, 2000). A study in cats sampling neurons in the different MGB anatomical subdivisions (Villa et al., 1991) revealed that the increases in signal-to-noise ratio (evoked divided by spontaneous firing rate) often result from a larger decrease in spontaneous than in evoked activity. Subsequent studies using pharmacological inactivation of auditory cortex by muscimol (a long-lasting $GABA_A$ agonist) or lidocaine (a local anesthetic acting on sodium channels), have reported that cortical inactivation reduced auditory responses in the ventral tonotopic lemniscal division of MGB (MGv) and in the inferior colliculus with a larger (60 vs. 34%) and faster (11 vs. 31 min) reduction for thalamic neurons than for collicular neurons (mustached bat, Zhang and Suga, 1997; Zhang et al., 1997). The effects of stimulating or blocking the activity of the auditory cortex while recording collicular neurons have been studied in different species. For example, Syka and Popelar (1984) in rats showed that most IC neurons, mainly located in the dorsal and caudal IC, reacted with a short excitation (3–15 ms) followed by inhibition lasting 30–150 ms or just inhibition after electrical stimulation of the auditory cortex (bipolar electrodes, single pulses, duration 0.2 ms, current 0.2–1.5 mA). Similar approach was used by Torterolo et al. (1998) with electrical stimulation in the guinea pig auditory cortex while recordings were performed in the IC neurons, observing differential effects on spontaneous and driven activity and different latencies depending on whether the recording was ipsilateral or contralateral to the stimulated cortex. Jen et al. (1998) recorded neurons in the CNIC of the big brown bat while blocking with lidocaine or electrically stimulating the auditory cortex. They showed corticofugal facilitation or inhibition, with longer latencies with inhibition. The cortical effect was most effective when it was combined with sounds of low intensity. The effects of phasic electrical stimulation of auditory cortex have pointed out the view that cortico-thalamic projections have an excitatory influence on thalamic activity. In guinea pigs, auditory cortex stimulation facilitated tone-evoked responses for more than 2/3 of the MGv neurons, especially when the BFs of the cortical and thalamic recordings were similar (He et al., 2002). Surprisingly, a similar cortical activation tended to induce inhibitory effects in the non-lemniscal divisions of the auditory thalamus (He, 2003), potentially due to the activation of GABAergic neurons from the thalamic reticular nucleus (Cotillon and Edeline, 2000) or from the IC (cats: Winer et al., 1996). Subsequent intracellular studies have confirmed this differential effect: depolarizations of MGB neurons in guinea pigs were only observed in the lemniscal division whereas hyperpolarizations were only observed in non-lemniscal MGB neurons (Yu et al., 2004). These changes in membrane polarizations contribute to a differential change in the acoustic responses of MGB cells (Xiong et al., 2004). In addition, they also pointed out that stimulation of the auditory cortex can modulate evoked responses in the auditory sector of the reticular nucleus and also promote a more tonic mode of discharge (Xu et al., 2007). It was speculated that the systematic selectivity of facilitation and inhibition over the lemniscal and non-lemniscal MGB is related to the attention shift within the auditory modality and across the sensory modalities (Yu et al., 2004).

The techniques used in these initial studies had obvious limitations. Besides the risks of non-specific effects (such as lowering the blood temperature during cortical cooling), the main consequence of global inactivation of the whole auditory cortex is removing its input onto corticofugal targets, including MGB and IC cells, but also onto higher cortical areas. Likewise, cortical electrical stimulation can trigger neuronal discharge in subcortical cells by both orthodromic and antidromic activation. In addition, global electrical activation or chemical inactivation obviously affects all descending projections originating from the auditory cortex, not only those reaching the subcortical structure under investigation (the MGB or the inferior colliculus). To circumvent these limitations, optogenetic tools have been used in most recent studies, to transiently silence, or activate, auditory cortex neurons in anesthetized and awake animal.

Modulation of Cortical Projections by Optogenetic Techniques

As described in the first part of this review, some studies have suggested that there was a difference in neuronal adaptation to noise between cortical and subcortical structures (Rabinowitz et al., 2013). A more recent study in mice (Lohse et al., 2020) has reported that the contrast gain control was robust in A1, MGv and CNIC. In these experiments, the degree

of adaptation to high (40 dB) or low (20 dB) contrast to dynamic random chords (DRC) was evaluated in MGv and CNIC during the silencing of cortical neurons (by activating inhibitory GABA interneurons). The contrast gain control was unchanged during cortical silencing both in anesthetized and awake mice at collicular level and in anesthetized animals at thalamic level, which clearly points out that subcortical neurons can exhibit contrast adaptation via intrinsic, cortical-independent mechanisms. Interestingly, cortical silencing had no effect on the shape of the spectro-temporal receptive fields (STRFs, i.e., BF value, spectral and temporal bandwidth, value of the largest weight in the kernel) both in MGv and CNIC. When the cortex was silenced, it is also interesting to note that (i) the reliability of responses to DRC was even increased in the MGv and in the CNIC of awake mice and that (ii) subcortical neurons were better described by a linear model than when the cortex was normally operating, as if the cortical inputs decrease the reliability and the linearity of MGv and CNIC neurons. Interestingly, in anesthetized or awake passively listening animals, the corticofugal projections did not contribute to the contrast adaptation observed in the MGv and CNIC.

However, and as it is the case with cortical cooling, one can consider that silencing the whole auditory cortex does not mimic a physiological situation. The corticofugal projections are topographically organized: Anterograde tracing studies have shown that the location of the terminal fields in the CNIC varies topographically with the location of the injection sites in A1 (rats: Saldaña et al., 1996; gerbils: Bajo and Moore, 2005; ferrets: Bajo et al., 2007, **Figure 3A**). Injecting tracers at two locations in ferret A1, where neurons were tuned to different frequencies, produced two distinct bands of labeling in the CNIC, suggesting that the A1-CNIC projection links neurons in both structures with similar frequency tuning (Bajo et al., 2007). This has been confirmed physiologically in the guinea pig by positioning multi-site probes along the tonotopic axes of A1 and the CNIC (Lim and Anderson, 2007). Thus, the activation or inactivation of projections coming from specific cortical frequency bands would shed light about the direct action of A1 neurons on CNIC cells sharing similar tuning properties and reaching similar frequency regions in MGv or CNIC. In addition and regarding A1-MGv projections, Homma et al. (2017) in ferrets have demonstrated mistuning sensitivity in MGv neurons and that feedback from A1 to MGv is required for the normal ability of animals to detect a mistuned harmonic within a complex sound. These studies confirmed the point-to-point connections between the auditory cortex and the subcortical auditory structures.

In a recent experiment in mice using a combination of cortico-anterograde and collicular retrograde viral transfection, it was possible to achieve viral specific transfection of only cortico-collicular neurons (Blackwell et al., 2020). This combination of techniques ensures that only neurons expressing *Cre* recombinase in the auditory cortex would express ChannelRhodopsine2 (ChR2) or a hyperpolarizing opsin (ArchT) in the auditory cortex. Opsins were expressed in AC-IC projecting neurons, and shining light over AC would directly activate, or suppress, only the cortico-collicular feedback projections (Blackwell et al., 2020). ChR2 activation of AC-IC neurons resulted in increasing spontaneous activity in IC neurons with decrease driven activity to pure tones and clicks, but with particularly small effects on magnitude. ArchT silencing of the same pathway has no effect on evoked activity on IC neurons. Both optogenetic manipulations suggest that cortico-collicular feedback does not provide strong modulation on passive listening mice under anesthesia or awake conditions. Consistent with the known cortico-collicular projections, the effects were observed mainly for cells located in the dorsal cortex of the IC (DCIC), not in CNIC. The small reduction in evoked response did not affect the selectivity of IC neurons and did not change the noise correlations during spontaneous and evoked activity. In the same experiment, the authors have tried to determine whether modulating the cortical inhibitory interneurons can change collicular responses (Blackwell et al., 2020). Whereas modulating parvalbumin (PV) interneurons had no effect on spontaneous and tone-evoked activity in IC, suppressing the activity of somatostatin (SST) interneurons increased spontaneous activity in IC. Altogether, this careful study performed both in anesthetized and awake, but passively listening, mice has revealed very little effect of the cortico-collicular projections in such listening conditions.

The main question that can be raised is whether the cortico-feedback projections only exert a strong influence behaving, actively listening, animals. To answer this question, it was necessary to train animals in behavioral tasks and determine the impact of temporary suppression of cortical feedback on behavioral performance.

Inactivating Specific Auditory Cortex Projections During Challenging Behavioral Tasks

One of the earliest studies that explored the behavioral consequences of suppressing the corticofugal inputs used Elvax implants to release chronically Muscimol, a $GABA_a$ agonist (Smith et al., 2004). Ferrets bilaterally implanted with muscimol-Elvax over A1 were trained in a sound localization task with short (40 ms) or long (100–1,000 ms) tone bursts. The implanted animals initially displayed lower correct sound localization during the first sessions, but they improved over time and finally reached the same performance as the control animals. Comparing the silencing of primary and non-primary cortical areas (or making lesions of these areas) induced modest but significant deficits in sound localization and pointed out that the largest deficits were when silencing primary auditory cortex (Nodal et al., 2010, 2012).

In such experiments, the global silencing of the cortex was suppressing all cortical activity not only the feedback to the subcortical structures. To address the question of how cortico-collicular projections impact behavioral responses, two different techniques have been used in the same animal model. First, Bajo et al. (2010, **Figure 3B**) have used a chromophore-targeted neuronal degeneration technique to investigate the behavioral consequences of selectively eliminating layer V neurons projecting from primary auditory cortical areas to the inferior colliculus. This approach resulted in a loss of about

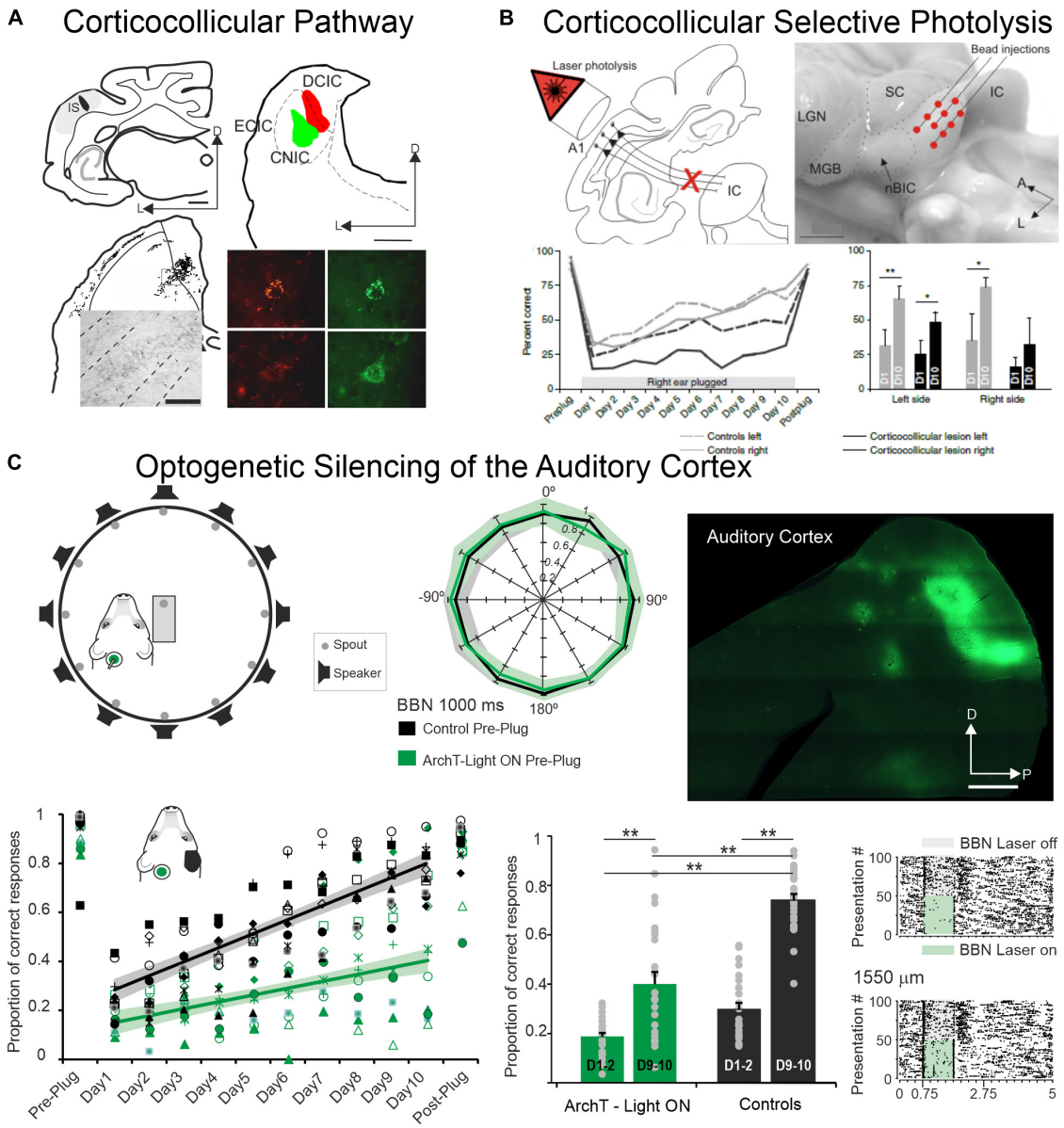

FIGURE 3 | The auditory cortex and the corticocollicular projection are essential for experience-dependent plasticity in spatial hearing. **(A)** Anterograde **(top left)** and retrograde **(top right)** tracer injections in the ferret the auditory cortex and in the Inferior Colliculus, respectively, reveal strong corticocollicular projection with terminal labeled fields in the three IC subdivisions **(bottom left)** after Fluororuby injection in A1 and retrogradely labeled cells in A1 **(bottom right)** after green and red fluorescent retrobead injections in the inferior colliculus. Modified from Bajo et al. (2007). **(B)** Chromophore-targeted laser photolysis of the corticocollicular pathway prevents learning-induced auditory plasticity. Corticocollicular layer V neurons were ablated using an infrared laser light **(top left)** following retrogradely neural labeling after microbead injections in the IC **(top right)**. Percent of correct responses in a sound localization task plotted against days of training including 10 days with unilateral right earplug **(bottom left)**. Data were grouped by left (dashed lines) and right (continuous lines) sound locations with control cases in gray and corticocollicular cases in black. In the bottom right, the mean and SD scores on the first (D1) and tenth day (D10) of monaural earplug are shown. Modified from Bajo et al. (2010). **(C)** Optogenetic silencing of the auditory cortex prevents earplug adaptation but not normal sound localization. Diagram shows the floor plan of the behavioral chamber **(top left)** and sound localization performance (proportion of correct responses at each speaker location). Data from control cases are in black and cases where neural activity in left A1 was optogenetic silenced using ArchT expression and green light illumination during each stimulus presentation in green **(middle panel)**. Histological section of a flattened auditory cortex showing GFP immunofluorescence associated with ArchT expression **(top right)**. Proportion of correct scores averaged across all speaker locations achieved by each animal in the control and A1 silenced groups (preplug session, 10 days with right earplug, and postplug) **(bottom left)**. Proportion of correct responses for the first and last 2 days of monaural occlusion **(middle panel)**. Examples of neural optogenetic suppression in A1 are shown in the bottom right panel. Neural responses driven by broadband stimulation [gray rectangles or combined with laser illumination (green rectangles)]. Modified from Bajo et al. (2019). $^{*}P < 0.05$, $^{**}P < 0.01$. Scale bars = 1 mm in **(A,C)**, 2 mm in **(B)**. A, anterior; A1, primary auditory cortex; BBN, broadband noise; CNIC, central nucleus of the inferior colliculus; D, dorsal; DCIC, dorsal cortex of the inferior colliculus; ECIC, external cortex of the inferior colliculus; HP, hippocampus; IC, inferior colliculus; IS, injection site; L, lateral; LGN, lateral geniculate nucleus; MGB, medial geniculate body; nBIC, nucleus of the brachium of the inferior colliculus; P, posterior; SC, superior colliculus.

two-thirds of the layer V A1 neurons that project to the IC, without affecting those in surrounding cortical areas or different cortical layers. Most cortico-collicular axons target the ipsilateral IC, so this approach allowed assessing the effects of removing descending axons on one side of the brain, although cross-projections comprise 15% of the cortico-collicular axons that were not eliminated. The behavioral results clearly indicate that ablation of the auditory cortico-collicular pathway from one hemisphere did not affect sound localization, as measured by either the initial orienting response to the sound or the subsequent selection of sound-source location. An interesting challenge was whether the lesioned animals would be able to localize sounds in altered conditions of sound localization such as the one occurring when one ear is occluded and, therefore when the values of binaural cues used for sound localization change (this task was initially described in Kacelnik et al., 2006). While control animals recover their ability to localize sounds accurately with training, despite the continued presence of a plug in one ear, this was not the case in ferrets in which the cortico-collicular projection had been largely removed (**Figure 3B**), suggesting that descending pathways are essential for recalibration of the brain's representation of auditory space. This learning deficit was most pronounced in the hemifield contralateral to the lesioned pathway, implying that corticofugal modulation of each IC mediates plasticity in the opposite hemifield (Bajo et al., 2010). Thus, one function of the auditory cortex in spatial hearing is to provide signals that are transmitted via descending cortical pathways to bring about experience-driven changes in localization.

Second, silencing auditory cortex neurons (by light stimulation of neurons expressing the proton pump ArchT) during sound presentations in an azimuthal sound-localization task did not impair the initial animals' behavioral performance (Bajo et al., 2019, **Figure 3C**): performance of control animals and the animals in which each stimulus presentation was paired with optogenetic silencing of A1 neurons localized broadband noise bursts was equally similar (Figure 1 in Bajo et al., 2019). When the animals were trained to re-learn the sound-localization task after unilateral ear occlusion (after plugging one ear), there was a massive drop in performance both in controls and in animals with optogenetic control of A1. Nonetheless, across 10 days of training to perform the task with monaural occlusion (note that plugging one ear change the values of the binaural cues but do not eliminate binaural cues), the control animals considerably improved their performance which was not the case for the animals for which A1 was silenced during each trial during sound delivery (**Figure 4** in Bajo et al., 2019, **Figure 3C**). Thus, suppressing auditory cortex activity did not prevent the animal to normally localized sounds, but impaired the ability to adapt to a unilateral earplug.

An additional surprising finding was observed when the same ear was occluded for a second time in control animals that had previously adapted to the unilateral hearing loss. A much smaller initial deficit was observed when the ear was replugged than when the animals first experienced an earplug. Furthermore, most of the control ferrets achieved their maximum score by ∼day 5 and remained at around that level until the end of the second period of monaural occlusion. In contrast, the ArchT animals (that had previously shown impaired adaptation when cortical activity was suppressed) re-tested with the occluded ear but *without silencing*

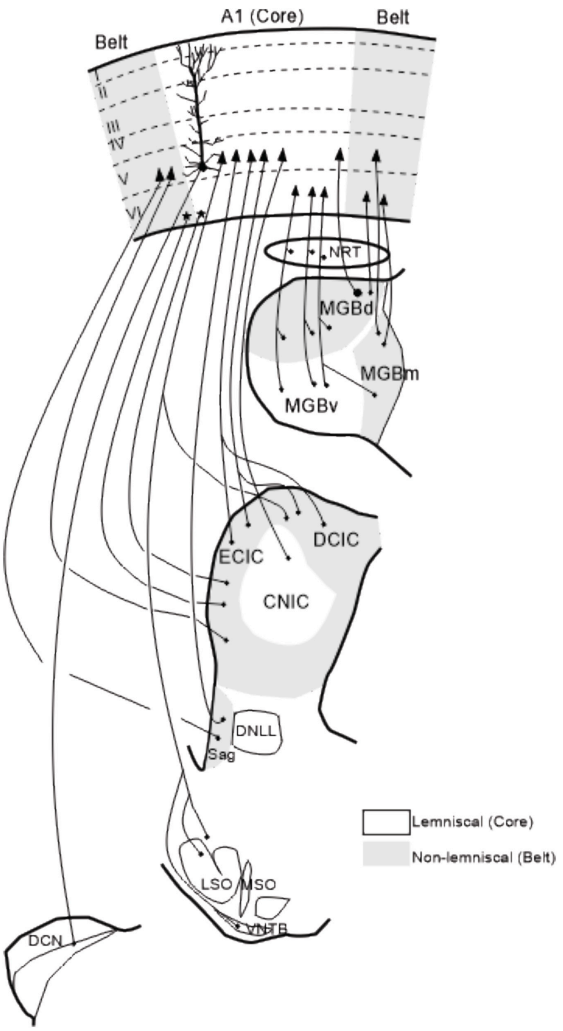

FIGURE 4 | Summary of corticofugal projections to auditory nuclei. Corticofugal projections to the nucleus reticularis thalamic, auditory thalamus and Sagulum are ipsilateral whereas projections to the inferior colliculus, superior olivary complex and cochlear nucleus are bilateral although predominantly from the same side. Modified from Winer (1992) and Malmierca et al. (2015). The details of these corticofugal projections are described in the following papers: Beyerl (1978); Adams (1980); Druga and Syka (1984a,b); Faye-Lund (1985); Feliciano and Potashner (1995); Saldaña et al. (1996); Weedman and Ryugo (1996); Beneyto et al. (1998); Winer et al. (1998); Winer et al. (2001); Jacomme et al. (2003); Bajo and Moore (2005); Meltzer and Ryugo (2006); Bajo et al. (2007); Coomes Peterson and Schofield (2007); Schofield (2009); Saldaña (2015); Lesicko and Llano (2017). A1, primary auditory cortex; CNIC, central nucleus of the inferior colliculus; DCN, dorsal cochlear nucleus; DCIC, dorsal cortex of the inferior colliculus; DNLL, dorsal nucleus of the lateral lemniscus; ECIC, external cortex of the inferior colliculus; LSO, lateral superior olive, MGBd, dorsal division of the medial geniculate body; MGBm, medial division of MGB; MGBv, ventral division of MGB; MSO, medial superior olive; NRT, nucleus reticularis of the thalamus; Sag, nucleus sagulum; VNTB, ventral nucleus of the trapezoid body; I–VI, cortical layers 1–6.

the auditory cortex did not show a better adaptation than during the first earplug. Despite a normal activity in the auditory cortex, these animals adapted at the same rate as that observed during the first period of monaural deprivation when A1 was inactivated, and significantly more slowly than the control animals during their first period of monaural occlusion. Thus, optogenetic suppression of cortical activity not only impairs auditory spatial learning, but also results in less effective adaptation when the active auditory cortex is subsequently challenged by monaural occlusion. When the auditory cortex was again inactivated on these animals, their performance was exactly the same as with the cortex intact, suggesting that the limited capacity of these animals to adapt to the second period of monaural occlusion no longer appears to be dependent on the activity of A1 (**Figure 6** of Bajo et al., 2019).

Both examples show the relevance of the auditory cortex and of the cortico-collicular projections in actively listening animals performing challenging behavior tasks.

DECIPHERING THE MECHANISMS UNDERLYING THE CORTICOFUGAL EFFECTS

Corticofugal projections are particular abundant in the auditory system (**Figure 4**; Winer, 2006). An important concept that has been proposed for understanding the functional role of corticofugal projections within the thalamo-cortical sensory systems is the distinction between "driver" and "modulator" inputs (Sherman and Guillery, 1998, 2002; Guillery and Sherman, 2002) which have been re-named Class 1 and Class 2 inputs (Lee and Sherman, 2010, 2011) based on the initial anatomical description by Guillery (1966). In the auditory system, this distinction leads to the possibility that the cortical afferents from A1 reaching MGv are modulatory inputs for the lemniscal (MGv) relay cells (review in Lee and Sherman, 2010, 2011). Note also that the impact of the cortical inputs can also be indirect via the thalamic reticular nucleus (TRN), which can have a stronger influence on the lemniscal division than on the non-lemniscal ones (cats: Crabtree, 1998; rats: Cotillon-Williams et al., 2008).

From the previous section, it seems that the crucial point that needs to be explored is how a cortical input projecting on IC cells (or MGv cells) which is, in some contexts, a modulator that modestly affects the functional properties of IC cells in awake passive animals (Blackwell et al., 2020) becomes a necessary input that can be used to drive the animal behavioral response (Bajo et al., 2010, 2019). In other words, what are the factors that, surprisingly, transform a potential weak and inefficient cortico-collicular input into a driving force that can guide the animal in its behavior? Could corticofugal projections act as drivers or modulators in a context dependent manner? The next question is how the subcortical networks are affected by cortical inputs depending on the difficulty of the task and the stability of those changes in time.

Here, we consider that in anesthetized animals and in awake animals that are not engaged in a behavioral challenging task, the corticofugal descending projections are only parts of the synaptic excitatory inputs reaching thalamic and collicular cells. In contrast, we would like to propose that the auditory corticofugal projections play an essential role during active listening associated to challenging behavioral tasks, under the dual control of neuromodulatory systems and the frontal cortical areas.

Neuromodulation in the Auditory Cortex

The most obvious factor that can change the way auditory stimuli are processed in awake animals between "passive" vs. "actively listening" conditions is the involvement of the neuromodulatory systems. Among them, the noradrenergic, dopaminergic and cholinergic systems have long been implicated in behavioral situations and cognitive functions (noradrenergic: Sara, 2009; dopaminergic: Seamans and Yang, 2004; Wise, 2004; Schultz, 2016; Ott and Nieder, 2019; cholinergic: Sarter et al., 2005; Lin et al., 2015). Two main properties should be considered about these neuromodulatory systems.

First, all brain nuclei at the origin of these neuromodulators are engaged, at different degrees, in cognitive functions. For example, neurons in the locus coeruleus (LC), the cortical source of noradrenaline (NA), are responsive to stimuli of any modality associated with reinforcements (Sara and Segal, 1991; Aston-Jones et al., 1997; Bouret and Sara, 2004). Dopaminergic neurons of the ventral tegmental area (VTA) are activated by rewards, and code for specific aspects of rewards such as their amount, probability of occurrence, subjective value, as well as to any reward-predicting stimuli, and their level of prediction of the reward occurrence (reviewed in Schultz, 2016). The cholinergic inputs arising from the basal forebrain (BF) area has long been involved in learning, acquired-stimulus salience and more generally in all situations of "attentional effort" (Sarter et al., 2006). In addition, experience dependent adaptation to the altered binaural cues was disrupted after the cortical cholinergic depletion in ferrets (Leach et al., 2013).

Needless to say, these three neuromodulatory systems do not work independently of each others, they all work in concert for controlling the state of cortical arousal and allowing cognitive performance. In fact, both in cortical and subcortical structures, non-synaptic interactions occurring at the presynaptic level are common and lead to subtle regulations of the excitatory and inhibitory transmission by a synergy between neuromodulators (reviewed in Vizi and Lábos, 1991; Vizi et al., 2010; Sperlágh and Vizi, 2011).

More importantly, these three neuromodulators drastically modify the processing of acoustic stimuli in the auditory cortex, and more generally, in the entire auditory system. For example, in guinea pigs, iontophoretic applications of NA increase the sharpness of tuning of auditory cortex neurons (Manunta and Edeline, 1997, 1998, 1999) and the neuronal discrimination performance between conspecific vocalizations (Gaucher and Edeline, 2015). Acetylcholine has a dual action on auditory cortex neurons. Whereas some effects were attributed to muscarinic receptors (mAChR; Guinea pigs: Metherate et al., 1990; Mice: Chen and Yan, 2007; Rats: Froemke et al., 2007), other studies proposed that the action of nicotinic receptors (nAChR) was prominent (Mice: Kawai et al., 2007; Rats:

Liang et al., 2006). In fact, activation of mAChRs tends to increase postsynaptic excitability while decreasing intracortical transmission via presynaptic receptors, whereas, in contrast, activation of nAChRs enhances thalamocortical transmission (reviewed in Edeline, 2003; Metherate, 2011). Only a few studies have described the dopaminergic modulation in the auditory cortex. In monkeys, it was shown that electrical stimulation of VTA modifies neuronal activity in the auditory cortex on two time scales: (i) effects on the time scale of tens to hundreds of milliseconds (Macaque monkeys: Mylius et al., 2015), and (ii) effect on the time scale of seconds and minutes that were reflected in the spontaneous and evoked activity (Huang et al., 2016). In gerbils, systemic administration of D1/D5 dopamine receptor agonists enhanced early infragranular auditory-evoked synaptic activity, prolonged auditory cortex activation, and more effectively recruited horizontal corticocortical networks during later phases of evoked activity (Happel et al., 2014). Note that neuromodulators alter auditory processing before the cortical level: Dopamine modulates the processing of unexpected auditory information as early as the inferior colliculus (Rats: Valdés-Baizabal et al., 2020), locus coeruleus activation alters thalamic and cortical responses to the same extent (Guinea pigs: Edeline et al., 2011), the pontomesencephalic cholinergic system modulates the activity of auditory thalamus and inferior colliculus (Woolf, 1991; Guinea pigs: Schofield et al., 2011), and NA modulates the response strength and the response latency as early as the cochlear nucleus (Mustached bat: Kössl and Vater, 1989) and by its action on the olivo-cochlear neurons can also modulate the compound action potential (Guinea pig: Mulders and Robertson, 2005a,b).

Although not historically considered major modulators of cortical processing, neuropeptides and neurohormones are now considered as such. For example, growing evidence suggests that oxytocin (OT) acts to enhance the salience of socially relevant sensory inputs and is important for parental behavior and social cognition. This peptide is synthesized in the paraventricular nucleus and supraoptic nucleus of the hypothalamus and binds to a G protein–coupled receptor with a single isoform (Gimpl and Fahrenholz, 2001). A series of studies have looked into the role of oxytocin in maternal behavior and in the processing of ultrasonic vocalization of pups when separated from the nest. Some studies have not found enhanced responses to pup calls between virgin and mother mice (Liu and Schreiner, 2007; Shepard et al., 2016; Royer et al., 2021). However, pharmacological application of oxytocin or optogenetic release of OT on the left auditory cortex (Marlin et al., 2015) reduced call-evoked inhibitory post-synaptic potentials (IPSCs) within seconds (**Figures 6A,B**, open and filled symbols respectively, and extended data **Figure 8** in Marlin et al., 2015), whereas the excitatory post-synaptic potentials (EPSCs) were gradually modified over minutes (**Figures 6A,B**, filled in Marlin et al., 2015). Therefore, oxytocin seems to rapidly disinhibit the auditory cortex (potentially similarly to ACh), suggesting that it can regulate attention and increase the salience of social stimuli. These results corroborate the effects of oxytocin in hippocampal slices (Rats: Owen et al., 2013).

Similar to oxytocin, orexins (Orexin A and B) are neuropeptides that profusely innervate the brain, including the deep layers of the neocortex (Marcus et al., 2001), and modulate the action of other classic neuromodulators (Peyron et al., 1998; but see Flores et al., 2015 for review). The orexin system is comprised of a small population of cells located mainly in the lateral hypothalamus. Orexins bind to specific receptors (OX1R and OX2R), associated with a Gq protein that activates the phospholipase C–protein kinase C pathway producing neural depolarization and increasing the membrane resistance by the closure of the K^+ conductance. Functions of the orexin system include the modulation of arousal and sleep–wake cycles, energy homeostasis, reward processing, stress and emotional behavior regulation (for example modulation of fear memory). Orexin might directly affect the auditory corticofugal pathways thanks to the specific expression of its receptors in layers V and VI (Rats: Marcus et al., 2001). In somatosensory and visual cortices, orexins induce functional changes in layer VIb neurons (Rats: Bayer et al., 2002). Layer VIb auditory neurons project to the inferior colliculus (Cats: Winer et al., 1998; Gerbils: Bajo and Moore, 2005; Guinea pigs: Schofield, 2009). In addition, the effect of the orexins might be indirectly mediated by the activation of the non-specific thalamocortical projections from the intrathalamic and midline nuclei (Bayer et al., 2002). Other indirect pathways might involve the medial prefrontal cortex (mPFC) cholinergic basal forebrain and locus coeruleus that show a great expression of orexin receptors (Marcus et al., 2001) and are capable as discussed above, of modulating auditory processing.

These studies indicate that in addition to the classical neuromodulators, oxytocin and orexins are also key actors to modulate the action of the cortical descending pathways.

Implications of the Frontal Areas in Attentional Processes During Active Auditory Listening

Attention is vital to achieve goals in constantly changing sensory environments. Frontal areas have long been suspected to play an important role in attentional processes. In primates, the auditory cortex projects and receives influence of higher order areas in the frontal cortex (Hackett et al., 1999; Romanski et al., 1999; Romanski and Averbeck, 2009). Over the last decades, electrophysiological recordings combined with behavioral tasks have demonstrated on one hand that, correlations of neuronal activity exist between the auditory cortex and frontal areas and, on the other hand, that there are also causal links between these two regions. Indeed, during tone detection tasks, Fritz et al. (2010) in ferrets showed that the activity of frontal cortex neurons was modulated by task events, but either by increasing or suppressing their firing rate to the target stimuli. In contrast, they lost responsiveness to identical stimuli presented passively, suggesting that frontal responses are tightly linked with the behavior. However, in these experiments, only a weak correlation between target response strength and task performance was observed. When the task was performed with visual cues, about one-third of the responsive frontal cells showed responses to both auditory and visual targets with similar responses to the two sensory modalities. The unimodal cells however presented different responses suggesting that some frontal cortex responses are modality specific. Interestingly, coherence analysis of local field potential (LFP) signals simultaneously recorded in A1

and frontal cortex showed that during active behavior, the synchronous activity between these areas is selectively enhanced when the target stimuli are presented but attenuated for responses to the reference sound. They argued that when an animal is engaged in a behavior, attention enhanced the synchronous activity between A1 and the frontal cortex.

To go further, Atiani et al. (2014) in the same animal model, compared responses obtained in A1, in two cortical belt areas and in dorsolateral frontal cortex during the same auditory discrimination task as Fritz et al. (2010). They showed that contrast enhancement between target and reference responses becomes more pronounced in frontal cortex than in auditory belt areas and than in A1. Thus, the reference responses are gradually suppressed as signals are transmitted through higher-order areas to frontal areas. In fact, recent analyses suggest that the neuronal responses became more categorical in higher cortical areas during task performance (Yin et al., 2020). Overall, these studies pointed out strong relationships between the activity in frontal and auditory cortex when an animal is engaged in an auditory discrimination task (Fritz et al., 2003, 2010; Atiani et al., 2014).

In primates, very few studies have investigated frontal cortical activity in various auditory behaviors to reveal the specific cognitive functions as decision making or reward value, associated with the network frontal cortex-auditory cortex. Tsunada et al. (2019) recorded neural activity from the ventrolateral prefrontal cortex (vlPFC) in two monkeys in a frequency discrimination task where they have to determine whether the tone bursts were predominantly "'low frequency"' or "'high frequency." They showed that post-decision vlPFC activity encodes the key features of the previous completed decision process that are used to generate the next one. Electrical microstimulation at vlPFC sites affected the monkeys' choices on the subsequent, but not the current, trial confirming that vlPFC activity is related to the encoding of the past trials and also informative in subsequent trials (for review, Banno et al., 2020).

Recently, Huang and Brosch (2020) recorded neuronal activity from vlFC in parallel with the neuronal activity from the auditory cortex of a single monkey performing two go-no go behavioral tasks requiring different audiomotor associations and using a sequence of two tones. Interestingly, they showed that, in the auditory cortex, the representations of the two tones were related to behavior. In contrast, in PFC, such a behavioral relevance was observed only for the first tone of the sequence. They thus promote the idea that the audiomotor representations in AC were more strongly related to behavior than those in PFC.

But does the activity in frontal areas provide enough excitatory inputs to drive auditory cortical neurons? Some studies used targeted-stimulation methodologies for demonstrating the relationships between neuronal activity in the frontal cortex and its effect in the auditory cortex in mice (Winkowski et al., 2013, 2018). First, the authors investigated the orbitofrontal cortex (OFC) stimulation on the neuronal activity in A1 using two-photon calcium imaging technique in mice (Winkowski et al., 2013). They found a diversity of effects, but often after pairing a particular frequency with the electric stimulation of OFC, the best frequency of A1 neurons in layers II/III changed with a response enhancement near the particular frequency used. Their results suggest that OFC activation could regulate neuronal activity within A1. Optogenetic activation of the mouse OFC in an area where neurons respond to sounds, activate A1 neurons and current source density (CSD) analysis revealed current sinks in layer I and layer IV, providing activation to both pyramidal cells and interneurons (Winkowski et al., 2018).

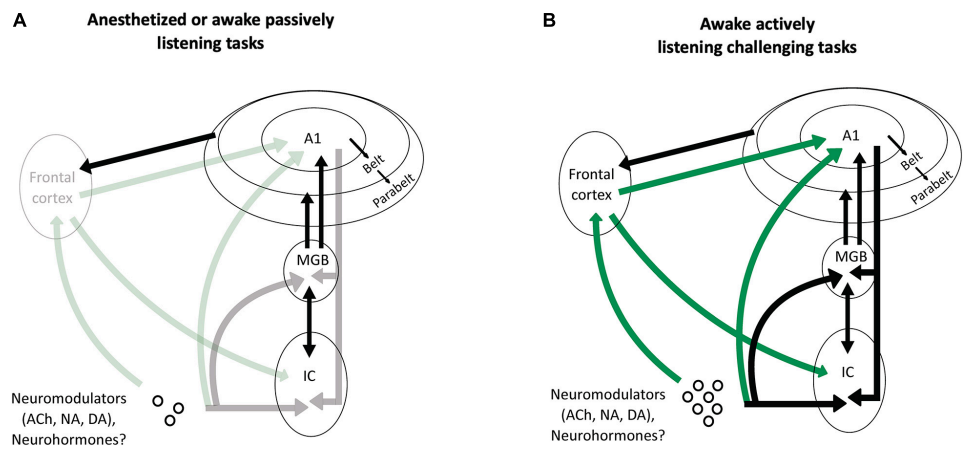

FIGURE 5 | Potential scenarios involving frontal projections and neuromodulators as key factors to increase the impact of the cortical descending projections on the thalamus and the inferior colliculus. **(A)** In anesthetized animals, or in awake passively listening animals performing easy auditory tasks, there is no or little activation of the frontal areas and of the neuromodulatory systems (symbolized by the shaded arrows). As a consequence, the descending projections from auditory cortex have a negligible effect on the processing of acoustic information by subcortical auditory structures (MGB and IC). **(B)** In awake actively listening animals performing challenging tasks, there is a strong activation of the frontal cortical areas and neuromodulatory systems (symbolized by the solid green lines). This activation sends a strong input to modulate the activity of the auditory cortex, which in turn, reshape the subcortical network (MGB and IC) thus allowing to perform successfully during this task.

Last, in a recent study, Olthof et al. (2019) described in adult rats, that the inferior colliculus, receives dense descending projections not only, from the auditory cortex, but also from the visual, somatosensory, motor, and prefrontal areas suggesting that the inferior colliculus can also integrate information coming from higher cortical areas.

POSSIBLE SCENARIOS

Based on the recent findings presented above, we propose that one of the fundamental roles of the frontal cortical areas and the neuromodulatory systems is to increase the efficacy that the cortical descending pathways exert on the subcortical structures when an animal is performing a challenging complex auditory task (**Figure 5**).

In anesthetized or awake passive conditions, or during basic auditory tasks, the descending projections from auditory cortex to subcortical structures are probably not necessary: in those cases, when the level of attention is absent (under anesthesia) or low (under passive listening), the frontal cortex and the neuromodulatory systems send no or little information to the cortical auditory areas. Under these conditions, the descending cortical inputs represent only a fraction of the excitatory inputs that subcortical neurons can use to build robust representations of the auditory scene. In contrast, when the task becomes challenging and the attentional level increases, the frontal areas and neuromodulatory systems are strongly activated (Humans: Berry et al., 2015; Du et al., 2016; Dimitrijevic et al., 2019; Monkeys: Lecas, 1995; reviewed in Peelle, 2018) and these inputs drastically change the activity of auditory cortex neurons. As a consequence, the descending cortical inputs to the subcortical auditory structures (MGB, IC, or even dorsal cochlear nucleus) send crucial information about the most adapted behavioral response needed to perform succesfully in these difficult conditions. In the case of the experiments discussed in this review, challenging conditions could be, for example, when an animal is engaged in discrimination tasks with noisy stimuli at very low SNRs, or during sound localization with an occluded ear.

Two possibilities can be envisioned for the emergence of the behavioral meaning at the subcortical level. Either the cortical descending inputs allow subcortical structures to generate more robust representations by plasticity mechanisms operating in the subcortical networks.

Alternatively, cortical input maximally account for the information already present at the subcortical level to perform the behaviorally challenging task. Collecting subcortical electrophysiological recordings during these challenging tasks with and without suppressing the descending cortical projections is probably the only way to determine which of these two assumptions is valid.

It is important to determine what are the relative contributions of the inputs from the frontal areas and from the neuromodulators. For example, data in humans suggest that subjects with a polymorphism of the choline transporter gene that is thought to limit choline transport capacity (Ile89Val variant of the choline transporter gene SLC5A7, rs1013940) do not show a robust activation of the right prefrontal cortex (Brodmann's areas 9) during challenging attentional tasks, whereas control subjects do (Berry et al., 2015). In addition, it is important to point out that the neuromodulators do not impact only the cortical level but also the subcortical structures, and that a single neuromodulator such as noradrenaline can influence auditory responses from cochlear nucleus (Mustached bats: Kössl and Vater, 1989) up to auditory cortex (Guinea pigs: Manunta and Edeline, 1997; Edeline, 1999; Gaucher and Edeline, 2015). Additionally, an important parameter that should be explored in the future is the timing of the network activation leading the animal to successful performance in a challenging task, and the stability of the network activation related to learning. A disruption in the network synchronization, or a delay in the activation of a key structure (as the frontal cortex, auditory cortex or in the release of some neuromodulators) could also contribute to behavior failure.

This dual control allows the auditory cortex to instruct subcortical structures about the meaning of each stimulus, its relationships with rewards, and the exact nature of the behavioral/motor response that need to be applied at the occurrence of a given stimulus in a particular environment. Although speculative, this scenario should be tested in future experiments.

AUTHOR CONTRIBUTIONS

SS and J-ME wrote the initial draft of the review. VMB and FRN revised the manuscript for important intellectual content and approved all aspects of the review. SS, VMB, and FRN designed the figures of the review. All authors contributed to the article and approved the submitted version.

FUNDING

FRN and VMB were supported by Wellcome Trust (WT108369/Z/2015/Z) and RNID funding (S52_Bajo). J-ME was supported by grants from the French Agence Nationale de la Recherche (ANR) (ANR-14-CE30-0019-01). SS was supported by the Fondation pour la Recherche Médicale (FRM) grant number ECO20160736099 and by the Entendre Foundation.

REFERENCES

Adams, J. C. (1980). Crossed and descending projections to the inferior colliculus. *Neurosci. Lett.* 19, 1–5. doi: 10.1016/0304-3940(80)90246-3

Aedo, C., Terreros, G., León, A., and Delano, P. H. (2016). The corticofugal effects of auditory cortex microstimulation on auditory nerve and superior olivary complex responses are mediated via Alpha-9 nicotinic receptor subunit. *PLoS One* 11:e0155991. doi: 10.1371/journal.pone.0155991

Anderson, R. A., Snyder, R. L., and Merzenich, M. M. (1980). The topographic organization of corticocollicular projections from physiologically identified loci in the AI, AII and anterior auditory cortical fields in the cat. *J. Comp. Neurol.* 191, 479–494. doi: 10.1002/cne.901910310

Aston-Jones, G., Rajkowski, J., and Kubiak, P. (1997). Conditioned responses of monkey locus coeruleus neurons anticipate acquisition of discriminative behavior in a vigilance task. *Neuroscience* 80, 697–715. doi: 10.1016/s0306-4522(97)00060-2

Atiani, S., David, S. V., Elgueda, D., Locastro, M., Radtke-Schuller, S., Shamma, S. A., et al. (2014). Emergent selectivity for task-relevant stimuli in higher-order auditory cortex. *Neuron* 82, 486–499. doi: 10.1016/j.neuron.2014.02.029

Aushana, Y., Souffi, S., Edeline, J. M., Lorenzi, C., and Huetz, C. (2018). Robust neuronal discrimination in primary auditory cortex despite degradations of spectro-temporal acoustic details: comparison between guinea pigs with normal hearing and mild age-related hearing loss. *J. Assoc. Res. Otolaryngol.* 19, 163–180. doi: 10.1007/s10162-017-0649-1

Bajo, V. M., and Moore, D. R. (2005). Descending projections from the auditory cortex to the inferior colliculus in the gerbil, Meriones unguiculatus. *J. Comp. Neurol.* 486, 101–116. doi: 10.1002/cne.20542

Bajo, V. M., Nodal, F. R., Bizley, J. K., Moore, D. R., and King, A. J. (2007). The ferret auditory cortex: descending projections to the inferior colliculus. *Cereb Cortex* 17, 475–491. doi: 10.1093/cercor/bhj164

Bajo, V. M., Nodal, F. R., Korn, C., Constantinescu, A. O., Mann, E. O., Boyden, E. S. 3rd, et al. (2019). Silencing cortical activity during sound-localization training impairs auditory perceptual learning. *Nat. Commun.* 10:3075.

Bajo, V. M., Nodal, F. R., Moore, D. R., and King, A. J. (2010). The descending corticocollicular pathway mediates learning-induced auditory plasticity. *Nat. Neurosci.* 13, 253–260. doi: 10.1038/nn.2466

Banno, T., Lestang, J. H., and Cohen, Y. E. (2020). Computational and neurophysiological principles underlying auditory perceptual decisions. *Curr. Opin. Physiol.* 18, 20–24. doi: 10.1016/j.cophys.2020.07.001

Bar-Yosef, O., and Nelken, I. (2007). The effects of background noise on the neural responses to natural sounds in cat primary auditory cortex. *Front. Comput. Neurosci.* 1:3. doi: 10.3389/neuro.10.003.2007

Bayer, L., Eggermann, E., Saint-Mleux, B., Machard, D., Jones, B. E., Muhlethaler, M., et al. (2002). Selective action of orexin (Hypocretin) on nonspecific thalamocortical projection neurons. *J. Neurosci.* 22, 7835–7839. doi: 10.1523/jneurosci.22-18-07835.2002

Beetz, M. J., García-Rosales, F., Kössl, M., and Hechavarría, J. C. (2018). Robustness of cortical and subcortical processing in the presence of natural masking sounds. *Sci. Rep.* 8:6863. doi: 10.1038/s41598-018-25241-x

Beneyto, M., Winer, J. A., Larue, D. T., and Prieto, J. J. (1998). Auditory connections and neurochemistry of the sagulum. *J. Comp. Neurol.* 401, 329–351. doi: 10.1002/(sici)1096-9861(19981123)401:3<329::aid-cne3>3.0.co;2-w

Berry, A. S., Blakely, R. D., Sarter, M., and Lustig, C. (2015). Cholinergic capacity mediates prefrontal engagement during challenges to attention: evidence from imaging genetics. *Neuroimage* 108, 386–395. doi: 10.1016/j.neuroimage.2014.12.036

Beyerl, B. D. (1978). Afferent projections to the central nucleus of the inferior colliculus in the rat. *Brain Res.* 145, 209–223. doi: 10.1016/0006-8993(78)90858-2

Bidelman, G. M., Davis, M. K., and Pridgen, M. H. (2018). Brainstem-cortical functional connectivity for speech is differentially challenged by noise and reverberation. *Hear Res.* 367, 149–160. doi: 10.1016/j.heares.2018.05.018

Bizley, J. K., Walker, K. M., Nodal, F. R., King, A. J., and Schnupp, J. W. (2013). Auditory cortex represents both pitch judgments and the corresponding acoustic cues. *Curr. Biol.* 23, 620–625. doi: 10.1016/j.cub.2013.03.003

Blackwell, J. M., Lesicko, A. M., Rao, W., De Biasi, M., and Geffen, M. N. (2020). Auditory cortex shapes sound responses in the inferior colliculus. *Elife* 9:e51890.

Bouret, S., and Sara, S. J. (2004). Reward expectation, orientation of attention and locus coeruleus-medial frontal cortex interplay during learning. *Eur. J. Neurosci.* 20, 791–802. doi: 10.1111/j.1460-9568.2004.03526.x

Brown, M. C., Mukerji, S., Drottar, M., Windsor, A. M., and Lee, D. J. (2013). Identification of inputs to olivocochlear neurons using transneuronal labeling with pseudorabies virus (PRV). *J. Assoc. Res. Otolaryngol. (JARO).* 14, 703–717. doi: 10.1007/s10162-013-0400-5

Carruthers, I. M., Laplagne, D. A., Jaegle, A., Briguglio, J. J., Mwilambwe-Tshilobo, L., Natan, R. G., et al. (2015). Emergence of invariant representation of vocalizations in the auditory cortex. *J. Neurophysiol.* 114, 2726–2740. doi: 10.1152/jn.00095.2015

Carruthers, I. M., Natan, R. G., and Geffen, M. N. (2013). Encoding of ultrasonic vocalizations in the auditory cortex. *J. Neurophysiol.* 109, 1912–1927. doi: 10.1152/jn.00483.2012

Chen, G., and Yan, J. (2007). Cholinergic modulation incorporated with a tone presentation induces frequency-specific threshold decreases in the auditory cortex of the mouse. *Eur. J. Neurosci.* 25, 1793–1803. doi: 10.1111/j.1460-9568.2007.05432.x

Coomes, D. L., and Schofield, B. R. (2004). Projections from the auditory cortex to the superior olivary complex in guinea pigs. *Eur. J. Neurosci.* 19, 2188–2200. doi: 10.1111/j.0953-816x.2004.03317.x

Coomes Peterson, D., and Schofield, B. R. (2007). Projections from auditory cortex contact ascending pathways that originate in the superior olive and inferior colliculus. *Hear Res.* 232, 67–77. doi: 10.1016/j.heares.2007.06.009

Cotillon, N., and Edeline, J. M. (2000). Tone-evoked oscillations in the rat auditory cortex result from interactions between the thalamus and reticular nucleus. *Eur. J. Neurosci.* 12, 3637–3650. doi: 10.1046/j.1460-9568.2000.00254.x

Cotillon-Williams, N., Huetz, C., Hennevin, E., and Edeline, J. M. (2008). Tonotopic control of auditory thalamus frequency tuning by reticular thalamic neurons. *J. Neurophysiol.* 99, 1137–1151. doi: 10.1152/jn.01159.2007

Crabtree, J. W. (1998). Organization in the auditory sector of the cat's thalamic reticular nucleus. *J. Comp. Neurol.* 390, 167–182. doi: 10.1002/(sici)1096-9861(19980112)390:2<167::aid-cne1>3.0.co;2-#

Dean, I., Harper, N. S., and McAlpine, D. (2005). Neural population coding of sound level adapts to stimulus statistics. *Nat. Neurosci.* 8, 1684–1689. doi: 10.1038/nn1541

Dean, I., Robinson, B. L., Harper, N. S., and McAlpine, D. (2008). Rapid neural adaptation to sound level statistics. *J. Neurosci.* 28, 6430–6438. doi: 10.1523/JNEUROSCI.0470-08.2008

Dimitrijevic, A., Smith, M. L., Kadis, D. S., and Moore, D. R. (2019). Neural indices of listening effort in noisy environments. *Sci. Rep.* 9:11278. doi: 10.1038/s41598-019-47643-1

Doucet, J. R., Rose, L., and Ryugo, D. K. (2002). The cellular origin of corticofugal projections to the superior olivary complex in the rat. *Brain Res.* 925, 28–41. doi: 10.1016/s0006-8993(01)03248-6

Druga, R., and Syka, J. (1984a). Ascending and descending projections to the inferior colliculus in the rat. *Physiol. Bohemoslov.* 33, 31–42.

Druga, R., and Syka, J. (1984b). Neocortical projections to the inferior colliculus in the rat. (An experimental study using anterograde degeneration techniques). *Physiol. Bohemoslov.* 33, 251–253.

Du, Y., Buchsbaum, B. R., Grady, C. L., and Alain, C. (2016). Increased activity in frontal motor cortex compensates impaired speech perception in older adults. *Nat. Commun.* 7:12241. doi: 10.1038/ncomms12241

Edeline, J. M. (1999). Learning-induced physiological plasticity in the thalamo-cortical sensory systems: a critical evaluation of receptive field plasticity, map changes and their potential mechanisms. *Prog. Neurobiol.* 57, 165–224. doi: 10.1016/s0301-0082(98)00042-2

Edeline, J. M. (2003). The thalamo-cortical auditory receptive fields: regulation by the states of vigilance, learning and the neuromodulatory systems. *Exp. Brain Res.* 153, 554–572. doi: 10.1007/s00221-003-1608-0

Edeline, J. M., Manunta, Y., and Hennevin, E. (2011). Induction of selective plasticity in the frequency tuning of auditory cortex and auditory thalamus neurons by locus coeruleus stimulation. *Hear Res.* 274, 75–84. doi: 10.1016/j.heares.2010.08.005

Edeline, J. M., Pham, P., and Weinberger, N. M. (1993). Rapid development of learning-induced receptive field plasticity in the auditory cortex. *Behav. Neurosci.* 107, 539–551. doi: 10.1037//0735-7044.107.4.539

Edeline, J. M., and Weinberger, N. M. (1993). Receptive field plasticity in the auditory cortex during frequency discrimination training: selective retuning independent of task difficulty. *Behav. Neurosci.* 107, 82–103. doi: 10.1037//0735-7044.107.1.82

Elgueda, P., and Delano, P. H. (2020). Corticofugal modulation of audition. *Curr. Opin. Physiol.* 18, 73–78. doi: 10.1016/j.cophys.2020.08.016

Faye-Lund, H. (1985). The neocortical projection to the inferior colliculus in the albino rat. *Anat. Embryol. (Berl.)* 173, 53–70. doi: 10.1007/BF00707304

Feliciano, M., and Potashner, S. J. (1995). Evidence for a glutamatergic pathway from the guinea pig auditory cortex to the inferior colliculus. *J. Neurochem.* 65, 1348–1357. doi: 10.1046/j.1471-4159.1995.65031348.x

Flores, A., Saravia, R., Maldonado, R., and Berrendero, F. (2015). Orexins and fear: implications for the treatment of anxiety disorders. *Trends Neurosci.* 38, 550–559. doi: 10.1016/j.tins.2015.06.005

Fritz, J., Shamma, S., Elhilali, M., and Klein, D. (2003). Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex. *Nat. Neurosci.* 6, 1216–1223. doi: 10.1038/nn1141

Fritz, J. B., David, S. V., Radtke-Schuller, S., Yin, P., and Shamma, S. A. (2010). Adaptive, behaviorally gated, persistent encoding of task-relevant auditory information in ferret frontal cortex. *Nat. Neurosci.* 13, 1011–1019. doi: 10.1038/nn.2598

Fritz, J. B., Elhilali, M., David, S. V., and Shamma, S. A. (2007). Does attention play a role in dynamic receptive field adaptation to changing acoustic salience in A1? *Hear Res.* 229, 186–203. doi: 10.1016/j.heares.2007.01.009

Fritz, J. B., Elhilali, M., and Shamma, S. A. (2005). Differential dynamic plasticity of A1 receptive fields during multiple spectral tasks. *J. Neurosci.* 25, 7623–7635. doi: 10.1523/JNEUROSCI.1318-05.2005

Froemke, R. C., Merzenich, M. M., and Schreiner, C. E. (2007). A synaptic memory trace for cortical receptive field plasticity. *Nature* 450, 425–429. doi: 10.1038/nature06289

Fuglsang, S. A., Dau, T., and Hjortkjær, J. (2017). Noise-robust cortical tracking of attended speech in real-world acoustic scenes. *Neuroimage* 156, 435–444. doi: 10.1016/j.neuroimage.2017.04.026

Gaucher, Q., and Edeline, J. M. (2015). Stimulus-specific effects of noradrenaline in auditory cortex: implications for the discrimination of communication sounds. *J. Physiol.* 593, 1003–1020. doi: 10.1113/jphysiol.2014.282855

Gimpl, G., and Fahrenholz, F. (2001). The oxytocin receptor system: structure, function, and regulation. *Physiol. Rev.* 81, 629–683. doi: 10.1152/physrev.2001.81.2.629

Guillery, R. W. (1966). A study of Golgi preparations from the dorsal lateral geniculate nucleus of the adult cat. *J. Comp. Neurol.* 128, 21–50. doi: 10.1002/cne.901280104

Guillery, R. W., and Sherman, S. M. (2002). Thalamic relay functions and their role in corticocortical communication: generalizations from the visual system. *Neuron* 33, 163–175. doi: 10.1016/s0896-6273(01)00582-7

Hackett, T. A., Stepniewska, I., and Kaas, J. H. (1999). Prefrontal connections of the parabelt auditory cortex in macaque monkeys. *Brain Res.* 817, 45–58. doi: 10.1016/s0006-8993(98)01182-2

Happel, M. F., Deliano, M., Handschuh, J., and Ohl, F. W. (2014). Dopamine-modulated recurrent corticoefferent feedback in primary sensory cortex promotes detection of behaviorally relevant stimuli. *J. Neurosci.* 34, 1234–1247. doi: 10.1523/JNEUROSCI.1990-13.2014

He, J. (2003). Corticofugal modulation on both ON and OFF responses in the nonlemniscal auditory thalamus of the guinea pig. *J. Neurophysiol.* 89, 367–381. doi: 10.1152/jn.00593.2002

He, J., Yu, Y. Q., Xiong, Y., Hashikawa, T., and Chan, Y. S. (2002). Modulatory effect of cortical activation on the lemniscal auditory thalamus of the Guinea pig. *J. Neurophysiol.* 88, 1040–1050. doi: 10.1152/jn.2002.88.2.1040

Hershenhoren, I., and Nelken, I. (2017). Detection of tones masked by fluctuating noise in rat auditory cortex. *Cereb Cortex.* 27, 5130–5143.

Homma, N. Y., Happel, M. F. K., Nodal, F. R., Ohl, F. W., King, A. J., and Bajo, V. M. A. (2017). Role for auditory corticothalamic feedback in the perception of complex sounds. *J. Neurosci.* 37, 6149–6161. doi: 10.1523/jneurosci.0397-17.2017

Hromádka, T., Deweese, M. R., and Zador, A. M. (2008). Sparse representation of sounds in the unanesthetized auditory cortex. *PLoS Biol.* 6:e16. doi: 10.1371/journal.pbio.0060016

Huang, Y., and Brosch, M. (2020). Associations between sounds and actions in primate prefrontal cortex. *Brain Res.* 1738:146775. doi: 10.1016/j.brainres.2020.146775

Huang, Y., Mylius, J., Scheich, H., and Brosch, M. (2016). Tonic effects of the dopaminergic ventral midbrain on the auditory cortex of awake macaque monkeys. *Brain Struct. Funct.* 221, 969–977. doi: 10.1007/s00429-014-0950-2

Jacomme, A. V., Nodal, F. R., Bajo, V. M., Manunta, Y., Edeline, J. M., Babalian, A., et al. (2003). The projection from auditory cortex to cochlear nucleus in guinea pigs: an *in vivo* anatomical and *in vitro* electrophysiological study. *Exp. Brain Res.* 153, 467–476. doi: 10.1007/s00221-003-1606-2

Jen, P. H., Chen, Q. C., and Sun, X. D. (1998). Corticofugal regulation of auditory sensitivity in the bat inferior colliculus. *J. Comp. Physiol.* 183, 683–697. doi: 10.1007/s003590050291

Kacelnik, O., Nodal, F. R., Parsons, C. H., and King, A. J. (2006). Training-induced plasticity of auditory localization in adult mammals. *PLoS Biol.* 4:e71. doi: 10.1371/journal.pbio.0040071

Kawai, H., Lazar, R., and Metherate, R. (2007). Nicotinic control of axon excitability regulates thalamocortical transmission. *Nat Neurosci.* 10, 1168–1175. doi: 10.1038/nn1956

Kössl, M., and Vater, M. (1989). Noradrenaline enhances temporal auditory contrast and neuronal timing precision in the cochlear nucleus of the mustached bat. *J. Neurosci.* 9, 4169–4178. doi: 10.1523/JNEUROSCI.09-12-04169.1989

Las, L., Stern, E. A., and Nelken, I. (2005). Representation of tone in fluctuating maskers in the ascending auditory system. *J. Neurosci.* 25, 1503–1513. doi: 10.1523/JNEUROSCI.4007-04.2005

Leach, N. D., Nodal, F. R., Cordery, P. M., King, A. J., and Bajo, V. M. (2013). Cortical cholinergic input is required for normal auditory perception and experience-dependent plasticity in adult ferrets. *J. Neurosci.* 33, 6659–6671. doi: 10.1523/JNEUROSCI.5039-12.2013

Lecas, J. C. (1995). Prefrontal neurones sensitive to increased visual attention in the monkey. *Neuroreport* 7, 305–309. doi: 10.1097/00001756-199512000-00073

Lee, C. C., and Middlebrooks, J. C. (2011). Auditory cortex spatial sensitivity sharpens during task performance. *Nat. Neurosci.* 14, 108–114. doi: 10.1038/nn.2713

Lee, C. C., and Sherman, S. M. (2010). Drivers and modulators in the central auditory pathways. *Front. Neurosci.* 4:79. doi: 10.3389/neuro.01.014.2010

Lee, C. C., and Sherman, S. M. (2011). On the classification of pathways in the auditory midbrain, thalamus, and cortex. *Hear Res.* 276, 79–87. doi: 10.1016/j.heares.2010.12.012

Lesica, N. A., and Grothe, B. (2008). Efficient temporal processing of naturalistic sounds. *PLoS One* 3:e1655. doi: 10.1371/journal.pone.0001655

Lesicko, A. M., and Llano, D. A. (2017). Impact of peripheral hearing loss on top-down auditory processing. *Hear Res.* 343, 4–13. doi: 10.1016/j.heares.2016.05.018

Liang, K., Poytress, B. S., Chen, Y., Leslie, F. M., Weinberger, N. M., and Metherate, R. (2006). Neonatal nicotine exposure impairs nicotinic enhancement of central auditory processing and auditory learning in adult rats. *Eur. J. Neurosci.* 24, 857–866. doi: 10.1111/j.1460-9568.2006.04945.x

Lim, H. H., and Anderson, D. J. (2007). Antidromic activation reveals tonotopically organized projections from primary auditory cortex to the central nucleus of the inferior colliculus in guinea pig. *J. Neurophysiol.* 97, 1413–1427. doi: 10.1152/jn.00384.2006

Lin, S. C., Brown, R. E., Hussain Shuler, M. G., Petersen, C. C., and Kepecs, A. (2015). Optogenetic dissection of the basal forebrain neuromodulatory control of cortical activation, plasticity, and cognition. *J. Neurosci.* 35, 13896–13903. doi: 10.1523/JNEUROSCI.2590-15.2015

Liu, R. C., and Schreiner, C. E. (2007). Auditory cortical detection and discrimination correlates with communicative significance. *PLoS Biol.* 5:e173. doi: 10.1371/journal.pbio.0050173

Lohse, M., Bajo, V. M., King, A. J., and Willmore, B. D. B. (2020). Neural circuits underlying auditory contrast gain control and their perceptual implications. *Nat. Commun.* 11:324.

Malmierca, M. S., Anderson, L. A., and Antunes, F. M. (2015). The cortical modulation of stimulus-specific adaptation in the auditory midbrain and thalamus: a potential neuronal correlate for predictive coding. *Front. Syst. Neurosci.* 9:19. doi: 10.3389/fnsys.2015.00019

Malmierca, M. S., and Ryugo, D. K. (2011). "Descending connections of auditory cortex to the midbrain and brain stem," in *The Auditory Cortex*, eds J. A. Winer and C. E. Schreiner (New York, NY: Springer), 189–208. doi: 10.1007/978-1-4419-0074-6_9

Manunta, Y., and Edeline, J. M. (1997). Effects of noradrenaline on frequency tuning of rat auditory cortex neurons. *Eur. J. Neurosci.* 9, 833–847. doi: 10.1111/j.1460-9568.1997.tb01433.x

Manunta, Y., and Edeline, J. M. (1998). Effects of noradrenaline on rate-level function of auditory cortex neurons: is there a "gating" effect of noradrenaline? *Exp. Brain Res.* 118, 361–372. doi: 10.1007/s002210050290

Manunta, Y., and Edeline, J. M. (1999). Effects of noradrenaline on frequency tuning of auditory cortex neurons during wakefulness and slow-wave sleep. *Eur. J. Neurosci.* 11, 2134–2150. doi: 10.1046/j.1460-9568.1999.00633.x

Marcus, J. N., Aschkenasi, C. J., Lee, C. E., Chemelli, R. M., Saper, C. B., Yanagisawa, M., et al. (2001). Differential expression of orexin receptors 1 and 2 in the rat brain. *J. Comp. Neurol.* 435, 6–25. doi: 10.1002/cne.1190

Marlin, B. J., Mitre, M., D'amour, J. A., Chao, M. V., and Froemke, R. C. (2015). Oxytocin enables maternal behaviour by balancing cortical inhibition. *Nature* 520, 499–504. doi: 10.1038/nature14402

Martin, E. M., West, M. F., and Bedenbaugh, P. H. (2004). Masking and scrambling in the auditory thalamus of awake rats by Gaussian and modulated noises. *Proc. Natl. Acad. Sci. U.S.A.* 101, 14961–14965. doi: 10.1073/pnas.0306879101

Meltzer, N. E., and Ryugo, D. K. (2006). Projections from auditory cortex to cochlear nucleus: a comparative analysis of rat and mouse. *Anat. Rec. A Discov. Mol. Cell. Evol. Biol.* 288, 397–408. doi: 10.1002/ar.a.20300

Mesgarani, N., David, S. V., Fritz, J. B., and Shamma, S. A. (2014). Mechanisms of noise robust representation of speech in primary auditory cortex. *Proc. Natl. Acad. Sci. U.S.A.* 111, 6792–6797. doi: 10.1073/pnas.1318017111

Metherate, R. (2011). Functional connectivity and cholinergic modulation in auditory cortex. *Neurosci. Biobehav. Rev.* 35, 2058–2063. doi: 10.1016/j.neubiorev.2010.11.010

Metherate, R., Ashe, J. H., and Weinberger, N. M. (1990). Acetylcholine modifies neuronal acoustic rate-level functions in guinea pig auditory cortex by an action at muscarinic receptors. *Synapse* 6, 364–368. doi: 10.1002/syn.890060409

Mulders, W. H., and Robertson, D. (2000). Evidence for direct cortical innervation of medial olivocochlear neurones in rats. *Hear Res.* 144, 65–72. doi: 10.1016/s0378-5955(00)00046-0

Mulders, W. H., and Robertson, D. (2005a). Catecholaminergic innervation of guinea pig superior olivary complex. *J. Chem. Neuroanat.* 30, 230–242. doi: 10.1016/j.jchemneu.2005.09.005

Mulders, W. H., and Robertson, D. (2005b). Noradrenergic modulation of brainstem nuclei alters cochlear neural output. *Hear Res.* 204, 147–155. doi: 10.1016/j.heares.2005.01.009

Mylius, J., Happel, M. F., Gorkin, A. G., Huang, Y., Scheich, H., and Brosch, M. (2015). Fast transmission from the dopaminergic ventral midbrain to the sensory cortex of awake primates. *Brain Struct. Funct.* 220, 3273–3294. doi: 10.1007/s00429-014-0855-0

Nagarajan, S. S., Cheung, S. W., Bedenbaugh, P., Beitel, R. E., Schreiner, C. E., and Merzenich, M. M. (2002). Representation of spectral and temporal envelope of twitter vocalizations in common marmoset primary auditory cortex. *J. Neurophysiol.* 87, 1723–1737. doi: 10.1152/jn.00632.2001

Narayan, R., Best, V., Ozmeral, E., McClaine, E., Dent, M., Shinn-Cunningham, B., et al. (2007). Cortical interference effects in the cocktail party problem. *Nat. Neurosci.* 10, 1601–1607. doi: 10.1038/nn2009

Nelken, I., Rotman, Y., and Bar Yosef, O. (1999). Responses of auditory-cortex neurons to structural features of natural sounds. *Nature* 397, 154–157. doi: 10.1038/16456

Ni, R., Bender, D. A., Shanechi, A. M., Gamble, J. R., and Barbour, D. L. (2017). Contextual effects of noise on vocalization encoding in primary auditory cortex. *J. Neurophysiol.* 117, 713–727. doi: 10.1152/jn.00476.2016

Nodal, F. R., Bajo, V. M., and King, A. J. (2012). Plasticity of spatial hearing: behavioural effects of cortical inactivation. *J. Physiol.* 590, 3965–3986. doi: 10.1113/jphysiol.2011.222828

Nodal, F. R., Kacelnik, O., Bajo, V. M., Bizley, J. K., Moore, D. R., and King, A. J. (2010). Lesions of the auditory cortex impair azimuthal sound localization and its recalibration in ferrets. *J. Neurophysiol.* 103, 1209–1225. doi: 10.1152/jn.00991.2009

Ohl, F. W., and Scheich, H. (2005). Learning-induced plasticity in animal and human auditory cortex. *Curr. Opin. Neurobiol.* 15, 470–477. doi: 10.1016/j.conb.2005.07.002

Olthof, B. M. J., Rees, A., and Gartside, S. E. (2019). Multiple nonauditory cortical regions innervate the auditory midbrain. *J. Neurosci.* 39, 8916–8928. doi: 10.1523/JNEUROSCI.1436-19.2019

Otazu, G. H., Tai, L. H., Yang, Y., and Zador, A. M. (2009). Engaging in an auditory task suppresses responses in auditory cortex. *Nat. Neurosci.* 12, 646–654. doi: 10.1038/nn.2306

Ott, T., and Nieder, A. (2019). Dopamine and cognitive control in prefrontal cortex. *Trends Cogn. Sci.* 23, 213–234. doi: 10.1016/j.tics.2018.12.006

Owen, S. F., Tuncdemir, S. N., Bader, P. L., Tirko, N. N., Fishell, G., and Tsien, R. W. (2013). Oxytocin enhances hippocampal spike transmission by modulating fast-spiking interneurons. *Nature* 500, 458–462. doi: 10.1038/nature12330

Peelle, J. E. (2018). Listening effort: how the cognitive consequences of acoustic challenge are reflected in brain and behavior. *Ear Hear* 39, 204–214. doi: 10.1097/AUD.0000000000000494

Peyron, C., Tighe, D. K., van den Pol, A. N., de Lecea, L., Heller, H. C., Sutcliffe, J. G., et al. (1998). Neurons containing hypocretin (orexin) project to multiple neuronal systems. *J. Neurosci.* 18, 9996–10015. doi: 10.1523/jneurosci.18-23-09996.1998

Rabinowitz, N. C., Willmore, B. D. B., King, A. J., and Schnupp, J. W. H. (2013). Constructing noise-invariant representations of sound in the auditory pathway. *PLoS Biol.* 11:e1001710. doi: 10.1371/journal.pbio.1001710

Romanski, L. M., and Averbeck, B. B. (2009). The primate cortical auditory system and neural representation of conspecific vocalizations. *Annu. Rev. Neurosci.* 32, 315–346. doi: 10.1146/annurev.neuro.051508.135431

Romanski, L. M., Bates, J. F., and Goldman-Rakic, P. S. (1999). Auditory belt and parabelt projections to the prefrontal cortex in the rhesus monkey. *J. Comp. Neurol.* 403, 141–157. doi: 10.1002/(sici)1096-9861(19990111)403:2<141::aid-cne1<3.0.co;2-v

Royer, J., Huetz, C., Occelli, F., Cancela, J. M., and Edeline, J. M. (2021). Enhanced discriminative abilities of auditory cortex neurons for pup calls despite reduced evoked responses in C57BL/6 mother mice. *Neuroscience* 453, 1–16. doi: 10.1016/j.neuroscience.2020.11.031

Ryugo, D. K., and Weinberger, N. M. (1976). Corticofugal modulation of the medial geniculate body. *Exp. Neurol.* 51, 377–391. doi: 10.1016/0014-4886(76)90262-4

Saldaña, E. (2015). All the way from the cortex: a review of auditory corticosubcollicular pathways. *Cerebellum* 14, 584–596. doi: 10.1007/s12311-015-0694-4

Saldaña, E., Feliciano, M., and Mugnaini, E. (1996). Distribution of descending projections from primary auditory neocortex to inferior colliculus mimics the topography of intracollicular projections. *J. Comp. Neurol.* 371, 15–40. doi: 10.1002/(sici)1096-9861(19960715)371:1<15::aid-cne2>3.0.co;2-o

Sara, S. J. (2009). The locus coeruleus and noradrenergic modulation of cognition. *Nat. Rev. Neurosci.* 10, 211–223. doi: 10.1038/nrn2573

Sara, S. J., and Segal, M. (1991). Plasticity of sensory responses of locus coeruleus neurons in the behaving rat: implications for cognition. *Prog. Brain Res.* 88, 571–585. doi: 10.1016/s0079-6123(08)63835-2

Sarter, M., Gehring, W. J., and Kozak, R. (2006). More attention must be paid: the neurobiology of attentional effort. *Brain Res. Rev.* 51, 145–160. doi: 10.1016/j.brainresrev.2005.11.002

Sarter, M., Hasselmo, M. E., Bruno, J. P., and Givens, B. (2005). Unraveling the attentional functions of cortical cholinergic inputs: interactions between signal-driven and cognitive modulation of signal detection. *Brain Res. Brain Res. Rev.* 48, 98–111. doi: 10.1016/j.brainresrev.2004.08.006

Schneider, D. M., and Woolley, S. M. N. (2013). Sparse and background-invariant coding of vocalizations in auditory scenes. *Neuron* 79, 141–152. doi: 10.1016/j.neuron.2013.04.038

Schofield, B. R. (2009). Projections to the inferior colliculus from layer VI cells of auditory cortex. *Neuroscience* 159, 246–258. doi: 10.1016/j.neuroscience.2008.11.013

Schofield, B. R., Motts, S. D., and Mellott, J. G. (2011). Cholinergic cells of the pontomesencephalic tegmentum: connections with auditory structures from cochlear nucleus to cortex. *Hear Res.* 279, 85–95. doi: 10.1016/j.heares.2010.12.019

Schultz, W. (2016). Dopamine reward prediction-error signalling: a two-component response. *Nat. Rev. Neurosci.* 17, 183–195. doi: 10.1038/nrn.2015.26

Seamans, J. K., and Yang, C. R. (2004). The principal features and mechanisms of dopamine modulation in the prefrontal cortex. *Prog. Neurobiol.* 74, 1–58. doi: 10.1016/j.pneurobio.2004.05.006

Shaheen, L. A., Slee, S. J., and David, S. V. (2020). Task engagement improves neural discriminability in the auditory midbrain of the marmoset monkey. *J. Neurosci.* 41, 284–297. doi: 10.1523/JNEUROSCI.1112-20.2020

Shepard, K. N., Chong, K. K., and Liu, R. C. (2016). Contrast enhancement without transient map expansion for species-specific vocalizations in core auditory cortex during learning. *eNeuro* 3:ENEURO.318–ENEURO.316. doi: 10.1523/ENEURO.0318-16.2016

Sherman, S. M., and Guillery, R. W. (1998). On the actions that one nerve cell can have on another: distinguishing "drivers" from "modulators". *Proc. Natl. Acad. Sci. U.S.A.* 95, 7121–7126. doi: 10.1073/pnas.95.12.7121

Sherman, S. M., and Guillery, R. W. (2002). The role of the thalamus in the flow of information to the cortex. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 357, 1695–1708. doi: 10.1098/rstb.2002.1161

Shetake, J. A., Wolf, J. T., Cheung, R. J., Engineer, C. T., Ram, S. K., and Kilgard, M. P. (2011). Cortical activity patterns predict robust speech discrimination ability in noise. *Eur. J. Neurosci.* 34, 1823–1838. doi: 10.1111/j.1460-9568.2011.07887.x

Slee, S. J., and David, S. V. (2015). Rapid task-related plasticity of spectrotemporal receptive fields in the auditory midbrain. *J. Neurosci.* 35, 13090–13102. doi: 10.1523/JNEUROSCI.1671-15.2015

Smith, A. L., Parsons, C. H., Lanyon, R. G., Bizley, J. K., Akerman, C. J., Baker, G. E., et al. (2004). An investigation of the role of auditory cortex in sound localization using muscimol-releasing Elvax. *Eur. J. Neurosci.* 19, 3059–3072. doi: 10.1111/j.0953-816X.2004.03379.x

Souffi, S., Lorenzi, C., Huetz, C., and Edeline, J. M. (2021). Robustness to noise in the auditory system: a distributed and predictable property. *eNeuro* 8:ENEURO.43–ENEURO.21. doi: 10.1523/ENEURO.0043-21.2021

Souffi, S., Lorenzi, C., Varnet, L., Huetz, C., and Edeline, J. M. (2020). Noise-sensitive but more precise subcortical representations coexist with robust cortical encoding of natural vocalizations. *J. Neurosci.* 40, 5228–5246. doi: 10.1523/JNEUROSCI.2731-19.2020

Sperlágh, B., and Vizi, E. S. (2011). The role of extracellular adenosine in chemical neurotransmission in the hippocampus and Basal Ganglia: pharmacological and clinical aspects. *Curr. Top Med. Chem.* 11, 1034–1046. doi: 10.2174/156802611795347564

Syka, J., and Popelar, J. (1984). Inferior colliculus in the rat: neuronal responses to stimulation of the auditory cortex. *Neurosci. Lett.* 51, 235–240. doi: 10.1016/0304-3940(84)90557-3

Terreros, G., and Delano, P. H. (2015). Corticofugal modulation of peripheral auditory responses. *Front. Syst. Neurosci.* 9:134. doi: 10.3389/fnsys.2015.00134

Thompson, A. M., and Thompson, G. C. (1993). Relationship of descending inferior colliculus projections to olivocochlear neurons. *J. Comp. Neurol.* 335, 402–412. doi: 10.1002/cne.903350309

Torterolo, P., Zurita, P., Pedemonte, M., and Velluti, R. A. (1998). Auditory cortical efferent actions upon inferior colliculus unitary activity in the guinea pig. *Neurosci. Lett.* 249, 172–176. doi: 10.1016/s0304-3940(98)00367-x

Town, S. M., Wood, K. C., and Bizley, J. K. (2018). Sound identity is represented robustly in auditory cortex during perceptual constancy. *Nat. Commun.* 9:4786. doi: 10.1038/s41467-018-07237-3

Tsunada, J., Cohen, Y., and Gold, J. I. (2019). Post-decision processing in primate prefrontal cortex influences subsequent choices on an auditory decision-making task. *Elife* 8:e46770. doi: 10.7554/eLife.46770

Valdés-Baizabal, C., Carbajal, G. V., Pérez-González, D., and Malmierca, M. S. (2020). Dopamine modulates subcortical responses to surprising sounds. *PLoS Biol.* 18:e3000744. doi: 10.1371/journal.pbio.3000744

Villa, A. E., Rouiller, E. M., Simm, G. M., Zurita, P., de Ribaupierre, Y., and de Ribaupierre, F. (1991). Corticofugal modulation of the information processing in the auditory thalamus of the cat. *Exp. Brain Res.* 86, 506–517. doi: 10.1007/BF00230524

Vizi, E. S., Fekete, A., Karoly, R., and Mike, A. (2010). Non-synaptic receptors and transporters involved in brain functions and targets of drug treatment. *Br. J. Pharmacol.* 160, 785–809. doi: 10.1111/j.1476-5381.2009.00624.x

Vizi, E. S., and Lábos, E. (1991). Non-synaptic interactions at presynaptic level. *Prog. Neurobiol.* 37, 145–163. doi: 10.1016/0301-0082(91)90025-v

Weedman, D. L., and Ryugo, D. K. (1996). Projections from auditory cortex to the cochlear nucleus in rats: synapses on granule cell dendrites. *J. Comp. Neurol.* 371, 311–324. doi: 10.1002/(SICI)1096-9861(19960722)371:2<311::AID-CNE10<3.0.CO;2-V

Weinberger, N. M. (2004). Specific long-term memory traces in primary auditory cortex. *Nat. Rev. Neurosci.* 5, 279–290. doi: 10.1038/nrn1366

Weinberger, N. M., and Diamond, D. M. (1987). Physiological plasticity in auditory cortex: rapid induction by learning. *Prog. Neurobiol.* 29, 1–55. doi: 10.1016/0301-0082(87)90014-1

Winer, J. A. (1992). "The functional architecture of the medial geniculate body and primary auditory cortex," in *Springer Handbook of Auditory Research. The Mammalian Auditory Pathway: Neuroanatomy*, Vol. 1, eds D. B. Webster, A. N. Popper, and R. R. Fay (New York, NY: Springer), 222–409. doi: 10.1007/978-1-4612-4416-5_6

Winer, J. A. (2006). Decoding the auditory corticofugal systems. *Hear Res.* 212, 1–8. doi: 10.1016/j.heares.2005.06.014

Winer, J. A., Diehl, J. J., and Larue, D. T. (2001). Projections of auditory cortex to the medial geniculate body of the cat. *J. Comp. Neurol.* 430, 27–55. doi: 10.1002/1096-9861(20010129)430:1<27::aid-cne1013>3.0.co;2-8

Winer, J. A., Larue, D. T., Diehl, J. J., and Hefti, B. J. (1998). Auditory cortical projections to the cat inferior colliculus. *J. Comp. Neurol.* 400, 147–174. doi: 10.1002/(sici)1096-9861(19981019)400:2<147::aid-cne1>3.0.co;2-9

Winer, J. A., and Lee, C. C. (2007). The distributed auditory cortex. *Hear Res.* 229, 3–13. doi: 10.1016/j.heares.2007.01.017

Winer, J. A., Saint Marie, R. L., Larue, D. T., and Oliver, D. L. (1996). GABAergic feedforward projections from the inferior colliculus to the medial geniculate body. *Proc. Natl. Acad. Sci. U.S.A.* 93, 8005–8010. doi: 10.1073/pnas.93.15.8005

Winkowski, D. E., Bandyopadhyay, S., Shamma, S. A., and Kanold, P. O. (2013). Frontal cortex activation causes rapid plasticity of auditory cortical processing. *J. Neurosci.* 33, 18134–18148. doi: 10.1523/JNEUROSCI.0180-13.2013

Winkowski, D. E., Nagode, D. A., Donaldson, K. J., Yin, P., Shamma, S. A., Fritz, J. B., et al. (2018). Orbitofrontal cortex neurons respond to sound and activate primary auditory cortex neurons. *Cereb Cortex* 28, 868–879. doi: 10.1093/cercor/bhw409

Wise, R. A. (2004). Dopamine, learning and motivation. *Nat. Rev. Neurosci.* 5, 483–494. doi: 10.1038/nrn1406

Wittekindt, A., Kaiser, J., and Abel, C. (2014). Attentional modulation of the inner ear: a combined otoacoustic emission and EEG study. *J. Neurosci.* 34, 9995–10002. doi: 10.1523/JNEUROSCI.4861-13.2014

Wood, K. C., Town, S. M., and Bizley, J. K. (2019). Neurons in primary auditory cortex represent sound source location in a cue-invariant manner. *Nat. Commun.* 10:3019. doi: 10.1038/s41467-019-10868-9

Woolf, N. J. (1991). Cholinergic systems in mammalian brain and spinal cord. *Prog. Neurobiol.* 37, 475–524. doi: 10.1016/0301-0082(91)90006-m

Xiong, Y., Yu, Y. Q., Chan, Y. S., and He, J. (2004). Effects of cortical stimulation on auditory-responsive thalamic neurones in anaesthetized guinea pigs. *J. Physiol.* 560(Pt 1), 207–217. doi: 10.1113/jphysiol.2004.067686

Xu, M., Liu, C. H., Xiong, Y., and He, J. (2007). Corticofugal modulation of the auditory thalamic reticular nucleus of the guinea pig. *J. Physiol.* 585(Pt 1), 15–28. doi: 10.1113/jphysiol.2007.142240

Yin, P., Strait, D. L., Radtke-Schuller, S., Fritz, J. B., and Shamma, S. A. (2020). Dynamics and hierarchical encoding of non-compact acoustic categories in auditory and frontal cortex. *Curr. Biol.* 30, 1649.e5–1663.e5. doi: 10.1016/j.cub.2020.02.047

Yu, Y. Q., Xiong, Y., Chan, Y. S., and He, J. (2004). Corticofugal gating of auditory information in the thalamus: an *in vivo* intracellular recording study. *J. Neurosci.* 24, 3060–3069. doi: 10.1523/jneurosci.4897-03.2004

Zhang, Y., and Suga, N. (1997). Corticofugal amplification of subcortical responses to single tone stimuli in the mustached bat. *J. Neurophysiol.* 78, 3489–3492. doi: 10.1152/jn.1997.78.6.3489

Zhang, Y., Suga, N., and Yan, J. (1997). Corticofugal modulation of frequency processing in bat auditory system. *Nature* 387, 900–903. doi: 10.1038/43180

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Souffi, Nodal, Bajo and Edeline. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Neuroscience

BRIEF RESEARCH REPORT
published: 16 August 2021
doi: 10.3389/fnins.2021.718230

Auditory Cortical Changes Precede Brainstem Changes During Rapid Implicit Learning: Evidence From Human EEG

Erika Skoe[1], Jennifer Krizman[2], Emily R. Spitzer[3] and Nina Kraus[2,4,5,6]*

[1] Department of Speech, Language and Hearing Sciences, Connecticut Institute for Brain and Cognitive Sciences, University of Connecticut, Storrs, CT, United States, [2] Auditory Neuroscience Laboratory, Department of Communication Sciences, Northwestern University, Evanston, IL, United States, [3] Department of Otolaryngology, Head and Neck Surgery, New York University Grossman School of Medicine, New York, NY, United States, [4] Department of Neurobiology and Physiology, Northwestern University, Evanston, IL, United States, [5] Department of Otolaryngology, Northwestern University, Evanston, IL, United States, [6] Institute for Neuroscience, Northwestern University, Evanston, IL, United States

The auditory system is sensitive to stimulus regularities such as frequently occurring sounds and sound combinations. Evidence of regularity detection can be seen in how neurons across the auditory network, from brainstem to cortex, respond to the statistical properties of the soundscape, and in the rapid learning of recurring patterns in their environment by children and adults. Although rapid auditory learning is presumed to involve functional changes to the auditory network, the chronology and directionality of changes are not well understood. To study the mechanisms by which this learning occurs, auditory brainstem and cortical activity was simultaneously recorded *via* electroencephalogram (EEG) while young adults listened to novel sound streams containing recurring patterns. Neurophysiological responses were compared between easier and harder learning conditions. Collectively, the behavioral and neurophysiological findings suggest that cortical and subcortical structures each provide distinct contributions to auditory pattern learning, but that cortical sensitivity to stimulus patterns likely precedes subcortical sensitivity.

Keywords: auditory system, corticofugal, online learning, frequency following response (FFR), statistical learning

OPEN ACCESS

Edited by:
Mari Tervaniemi, University of Helsinki, Finland

Reviewed by:
Manuel S. Malmierca, University of Salamanca, Spain
Marc Schönwiesner, Leipzig University, Germany

****Correspondence:***
Erika Skoe
erika.skoe@uconn.edu

Specialty section:
This article was submitted to Auditory Cognitive Neuroscience, a section of the journal Frontiers in Neuroscience

Received: *31 May 2021*
Accepted: *20 July 2021*
Published: *16 August 2021*

Citation:
Skoe E, Krizman J, Spitzer ER and Kraus N (2021) Auditory Cortical Changes Precede Brainstem Changes During Rapid Implicit Learning: Evidence From Human EEG. Front. Neurosci. 15:718230. doi: 10.3389/fnins.2021.718230

INTRODUCTION

Natural sound environments are rich with temporal and spectral patterns that repeat over different timescales. To extract these patterns, the brain must analyze the soundscape to learn about its statistical properties, including the probability that two sounds repeatedly co-occur. This analysis happens rapidly and often without conscious awareness. Evidence of rapid neural computations relating to predictive coding can be observed across the central auditory network, from brainstem to auditory cortex (Carbajal and Malmierca, 2018). Within the auditory system, brainstem and cortical structures also operate reciprocally through ascending and descending pathways (Winer, 2006). Through the descending corticofugal pathway, the auditory cortex can alter the input it receives, inducing short-term changes and long-term subcortical reorganization that either facilitate or

attenuate subcortical processing of specific stimulus features (Suga et al., 2002). The corticofugal system appears to play an important role in auditory learning (de Boer and Thornton, 2008; Bajo et al., 2010). However, how learning ultimately emerges from these network processes is poorly understood, and questions remain about the degree to which brainstem and cortical structures independently, or dependently, contribute to different learning stages.

Pattern learning ("statistical learning") is viewed as a general-purpose mechanism that underlies language and music learning (Saffran et al., 1997; Saffran, 2003; Misyak and Christiansen, 2012). Despite significant behavioral evidence of statistical learning, neurophysiological investigations of human auditory learning rarely examine this type of learning and when the neurophysiological correlates of short-term auditory learning have been investigated, they generally focus on cortical (Tremblay et al., 2001; Ben-David et al., 2011) or subcortical structures (Hornickel et al., 2012; Song et al., 2012) in isolation. Here we focus on both. To study the neural correlates of rapid pattern learning we coupled behavioral measures of learning with electroencephalogram (EEG) recordings, using an approach that allowed us to extract cortical and subcortical activity from the same EEG recording (Font-Alaminos et al., 2021). EEG was recorded while adult humans passively listened to continuous sequences comprised of eight musical tones (C4, D4, E4, F4, F#4, G4, G#4, and A4) ranging in fundamental frequency (F0) (262–440 Hz). Sequences were designed so that the transitional probability (TP) between tones (i.e., the probability that one tone followed another) was either randomized to create an "unpatterned" condition or fixed to create "patterned" conditions. Two patterned conditions were used in the experiment, with different participant groups receiving each. These patterned conditions were created by pairing the eight tones into four doublets and fixing the doublet TP at 100% (e.g., C4 always followed E4). For these two patterned conditions, the same eight tones were used but the doublet set did not intersect. The inter-stimulus interval was the same between the patterned and unpatterned conditions so that the doublets in the patterned conditions could only be detected by their TPs and not conspicuous breaks between doublets. Despite having the same short- and long-term TPs, one of the patterned sets was harder to learn (Skoe et al., 2015). By comparing these two patterned conditions of varying difficulty, we aimed to capture auditory system plasticity at different stages of the learning process, while preserving stimulus features like F0, TPs, and interstimulus interval.

In EEG recordings, activity from various neuronal sub-populations is pooled into a single waveform. By selective signal filtering, low-frequency cortical potentials (<30 Hz) like the P1-P2 response, can be separated from the higher-frequency phase-locked response characteristic of brainstem activity (Skoe and Kraus, 2010). Tones, such as those used here, elicit transient responses time-locked to the stimulus onset and tonic responses time-locked to the F0 ("frequency following response," FFR). P1 and P2 are transient responses arising ∼100 and ∼200 ms (respectively) after the stimulus onset. They are generated in or near the primary auditory cortex, with the P2 generators extending to secondary auditory cortex and auditory association areas (Picton and Hillyard, 1974; Ponton et al., 2002; Martin et al., 2008). P1 and P2 are both sensitive to learning effects (Tremblay and Kraus, 2002; Ben-David et al., 2011), however, they have different maturational time courses (Ponton et al., 2002; Sussman et al., 2008), suggestive of unique neural processes. While P1 and P2 are distinguishable experimentally and developmentally, their unique functional significance is poorly understood, in part because P1 is generally small in adults (at the long interstimulus intervals usually used for cortical AEPs, >0.5 s) and because P2 typically co-varies with N1 (Crowley and Colrain, 2004). Like onset responses, tonic responses like FFRs can be observed across the neuro-axis. For frequencies >200 Hz, cortical phase-locking is weak to non-existent and thus FFRs measured to the frequencies used here likely reflects predominantly brainstem sources (Coffey et al., 2019; White-Schwoch et al., 2021).

Our previous report on this dataset focused on the FFR (Skoe et al., 2015). We found that the FFR to the patterned condition was different (smaller) from the unpatterned (baseline) condition, but only for the easier and not the harder condition. We now follow up on this finding by extracting and analyzing the P1-P2 cortical responses, which were recorded simultaneously with the FFRs. The goal was to use the combination of cortical (P1-P2) and brainstem (FFR) responses to better understand possible top-down, cortically driven, effects on the FFR during a short-term learning paradigm. We used task difficulty as a window into the directionality, and possible chronology, over which learning takes place across the auditory network. For the easier and harder conditions, different amounts of learning took place over the same exposure, creating the experimental conditions for studying EEG from brainstem and cortex at different stages of learning: a more advanced stage for the easier condition and an earlier stage for the harder condition.

MATERIALS AND METHODS

Thirty-six young adults, ages 18–26, participated in the learning paradigm. Written informed consent was obtained from all, with experimental protocols approved by Northwestern University's Institutional Review Board. Before testing, participants were pseudo-randomly grouped into two groups (n = 18/group). An additional 18 participants were tested on a control condition to confirm that the FFR, P1, and P2 components did not change upon repeated presentation of the unpatterned condition. The three groups were age and gender matched. Groups were also statistically matched with respect to pure tone hearing thresholds in the 250–8kHz range, auditory brainstem response Wave V latency, IQ, auditory working memory, total years of musical training, and performance on a musical skills test. The details of this can be found in our earlier report on this dataset, see Skoe et al. (2015).

The study included three phases (**Figure 1A**): Phase 1: Baseline EEG, Phase 2: Learning phase, and Phase 3: Testing phase. During Phase 1, EEG was recorded to the unpatterned condition to establish baseline levels of activity. During Phase 2, EEG was recorded while participants listened to one of two different

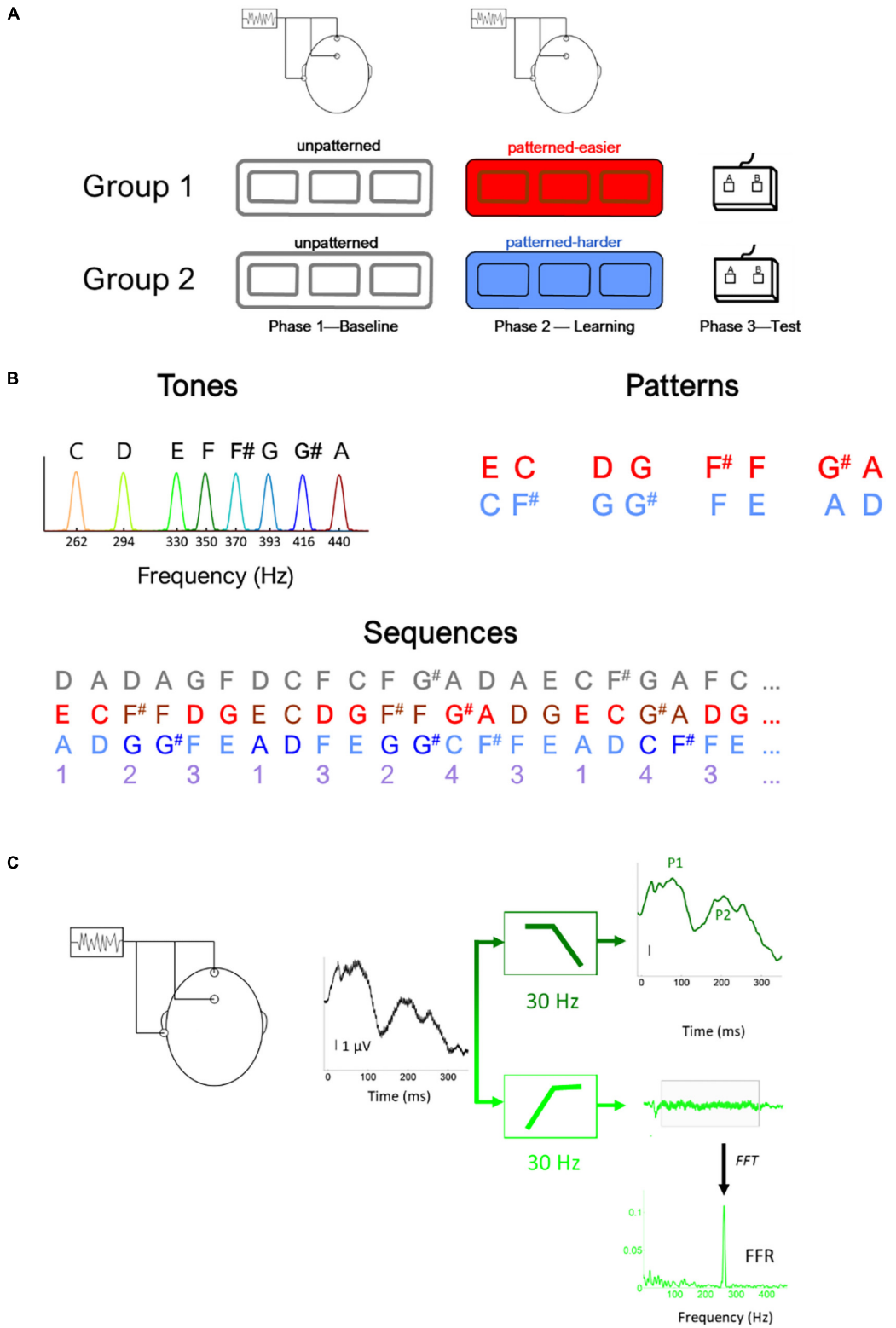

FIGURE 1 | Illustration of **(A)** experimental design, **(B)** stimulus (i.e., tones) and sequence characteristics, **(C)** filtering procedure to derive the brainstem (light green) and cortical (dark green) potentials, illustrated using the response to the lowest tone (262 Hz) **(C)**. In panel **(B)**, the gray numbers illustrated the unpatterned sequence and the purple numbers illustrate the generating sequence used to construct both the easier (red) and harder (blue) sequences. Using this generating sequence, 1 was replaced with EC for the easier sequence and AD for the harder sequence and 2 was replaced with F#F in the easier sequence and GG# in the harder sequence, etc. Thus, in addition to being composed of tones with the same fundamental frequencies and the same interstimulus interval between tones, the generating structure of the two pattern sequences had the same transitional probability distribution.

patterned conditions, as part of an implicit learning paradigm. In Phase 3, participants were tested behaviorally on whether they could recognize the patterns heard during Phase 2 and their performance served as behavioral measure of implicit learning. We opted to fix the presentation order of the unpatterned and patterned conditions rather than use an interleaved or counterbalanced order because of concerns that an interleaved order would interfere with implicit learning, and concerns that the patterned condition could influence the unpatterned (baseline) condition if presented first (Weiss et al., 2009).

At the outset of the Learning Phase, participants were told to listen carefully to the sounds and that they would later be tested on how well they remembered the sounds. To facilitate alertness while minimizing muscle movement, participants watched nature photos. Because statistical learning can be interrupted by a concurrent attention-demanding task (Toro and Trobalon, 2005), participants did not perform a photo-related or secondary task.

All three conditions (unpatterned and two patterned) were formed from the same eight complex tones (333 ms each), played at the same rate (2.7 tones/s, ISI = 37 ms). The F0s of the eight tones were 262, 294, 330, 350, 370, 393, 416, and 440 Hz with each tone mapping to a specific musical note (C4, D4, E4, F4, F#4, G4, G#4, and A4, respectively). The tones were triangle waves containing odd harmonics of the F0, where each successive harmonic diminished in amplitude by 1/X (X = harmonic number). These triangle wave stimuli were chosen because their natural clarinet-like sound quality is more pleasant to listen to than pure tones and because their spectral profile produces robust FFRs, especially when a small number of trials are used (Jeng et al., 2011; Tichko and Skoe, 2017).

Across the three conditions, each of the eight tones was presented with the same overall probability but different TPs. For the unpatterned condition, the TP was pseudo-randomized so that each tone had a roughly equal probability of being followed by another but could not follow itself (1/7 or 14.3% TP). See **Figure 1B** for an illustration of the first ∼8 s of each sequence. For the patterned sequences, tones were presented in pairs, and each pair was drawn from a pre-arranged set of four options, without direct repetition. The tone pair set was unique for each patterned sequence: Easier [EC, F#F, DG, G#A] or Harder [AD, G#G, FE, CF#]. The TP for each tone pair was 100%. During the Baseline and Learning Phases, the sequences were presented as 5-min blocks, with short 1-min breaks between blocks. Within each block, each tone was presented 100 times, for a total of 300 presentations. The experimental design included three blocks with the initial intention of studying the time-course of plasticity. While pilot testing suggested that 100 trials were sufficient to elicit robust FFRs, this did not bear out in the full study sample, where we found that 300 stimulus presentations needed to be averaged for the FFRs to the highest stimulus frequencies to be above the noise floor for many of the participants in the sample.

During the Testing Phase, participants were given a two-alternative forced-choice test in which each tone pair from the patterned sequence was presented with a foil pair, two sounds that were heard but never sequentially during the Learning Phase (Saffran et al., 1999; Abla et al., 2008). Participants were instructed to select the more familiar-sounding pair. Each pair was tested against four foils creating 16 comparisons, with each comparison tested once. The pairs forming one sequence were inverted to create the foils for the other. For example, the tone pair EC in one sequence was inverted to create the foil CE for the other sequence. Scores were converted to percent correct, with 50% representing chance. Pilot testing showed that the sequence comprised of [EC, F#F, DG, G#A] produced higher test scores [independent t test: $t(26) = 3.595$, $p = 0.001$], motivating us to label this sequence as "easier" and the other as "harder." This condition difference holds for the current dataset as well. In our previous publication of this dataset (Skoe et al., 2015), we reported that performance was at 61% correct for the easier condition compared to 53% for the harder condition. For the easier sequence, 15 of 18 (83%) participants performed above chance (i.e., 50%) compared to eight of 18 for the harder sequence. Within the easier sequence, F#F and G#A were more easily remembered than the other two tone pairs. While both sequences were novel and had similar TP distributions, the easier one was judged to be more musical by highly trained musicians tested during the pilot stage. This may account, at least in part, for why performance was different between the two sequences.

EEG Protocol

Electroencephalogram was recorded with an analog-to-digital rate of 20 kHz (SynAmps 2 amplifier, Neuroscan Acquire, Compumedics, Inc.). Three Ag-AgCl electrodes were placed on the scalp (non-inverting electrode at Cz, inverting electrode at A2, ground at FPz), with contact impedance <5 kilo ohms. Recordings were made in continuous mode with an online filter of 0.5–3,000 Hz and then were processed offline in Neuroscan Edit. An offline low-pass filter (<30 Hz, 12 dB/octave) isolated the cortical onset components of the recording. To extract the FFR, a 30–2,000 Hz (12 dB/octave) offline filter was applied (**Figure 1C**). After filtering, recordings were epoched with a window of −10 to 350 ms surrounding the onset presentation of each tone, and then baseline corrected to the mean voltage of the noise floor (−10 to 0 ms) before applying threshold-based artifact rejection criterion (FFR ± 35 μV, cortical response ± 100 μV). For each of the eight tones, 300 artifact-free trials were averaged for each participant (100/block) for the FFR and P1-P2 recordings.

The sustained component of the FFR (55–278 ms) was converted to the frequency domain using a fast Fourier transform. The FFR amplitude at the F0 of each tone was extracted following previously described procedures (Skoe et al., 2015). Our previous report showed that while some of the tone pairs were easier to remember than others in the easier condition, this was not reflected in the FFR of individual tones. Instead, the FFR effect emerged as a global reduction in amplitude across frequency. This prompted our focus on global and not tone-specific effects here. Because of this focus, the FFR-F0 amplitude was averaged across the eight tones to obtain a single value representing the response to the unpatterned and patterned sequences for each participant. Similarly, for the P1 and P2 analysis, P1 and P2 amplitudes were derived from the responses to the individual tones and then averaged to create a single value for each condition. For P1, the average amplitude was calculated

over 60–85 ms and for P2 the average amplitude was calculated over 160–220 ms. At the relatively fast rate of presentation used here (2.71 tones/s), P1 and P2 are the most prominent waveform components; N1, which is generally quite large in adults at slow presentation rates, is attenuated at faster rates (Sussman et al., 2008), and so was not analyzed here.

Statistical Analysis

Linear mixed-effects models were used to test for the effect of group and condition (patterned vs. unpatterned) (FFR amplitude, P1 amplitude, P2 amplitude), using subject ID as the random intercept. Statistical analyses were performed in MATLAB version R2019b using the function *fitlme* with the default covariance matrix structure (full covariance) and fit statistic method (maximum likelihood).

RESULTS

We first confirmed that repeating the unpatterned condition twice did not change the FFR, P1, or P2 components for the control group [*FFR* $t(34) = 1.24$, $p = 0.23$; *P1* $t(34) = 0.14$, $p = 0.89$; *P2* $t(34) = -1.82$, $p = 0.11$; example model formula = "*FFR*∼*1 + condition* + (*1*| *ID*)"]. Next, we confirmed the response to the unpatterned condition was not statistically different between the experimental groups who received the easier vs. harder patterned conditions [*FFR* $t(34) = 0.40$, $p = 0.69$; *P1* $t(34) = -0.61$, $p = 0.54$; *P2* $t(34) = -1.17$, $p = 0.25$; example model formula = "*FFR*∼*1 + Group* + (*1*| *ID*)"]. From there, we compared the two patterned conditions by testing whether the groups differed with respect to how much the amplitude changed between the unpatterned and patterned conditions [example model formula = "*FFR-change*∼*1 + Group* + (*1*| *ID*)"]. This analysis revealed that the easy and harder conditions differed for the FFR-change and P2-change measures, but not the P1-change [FFR $t(34) = -3.27$, $p < 0.01$; P1 $t(34) = -0.84$, $p = 0.40$; P2 $t(34) = -2.16$, $p = 0.04$] (**Figures 2**, **3**). Moreover, within-group comparisons between the patterned and unpatterned conditions showed that all three evoked responses were statistically smaller than baseline for the easier condition [FFR $t(34) = 3.45$, $p < 0.01$; P1 $t(34) = 2.75$, $p < 0.01$; P2 $t(34) = 2.44$, $p = 0.02$]. **Figure 3** shows the mean change across the eight tones for the three evoked responses for the easier and harder conditions. To illustrate that the reduction for the easier condition is global and no systematic tone-specific effects are apparent, the means are also graphed for the patterned and unpatterned conditions for each tone in **Figure 2**.

For the harder condition, *only* P1 was statistically different (smaller) from baseline [*FFR* $t(34) = -1.41$, $p = 0.17$; *P1* $t(34) = 2.42$, $p = 0.02$; *P2* $t(34) = -0.36$, $p = 0.72$] (**Figure 2**). This P1-effect for the harder condition was driven by the eight participants who performed above chance on the harder condition. For the 10 "non-learners" where performance was below chance, P1 in the patterned condition did not differ from P1 in the baseline condition [$t(18) = -0.14$, $p = 0.338$] but for the eight "learners" it did [$t(14) = 3.74$, $p < 0.01$]. Outside of showing different patterns for P1, the learners and non-learners did not differ with respect to FFR or P2 (both $p > 0.05$). Despite this group difference, the correlation between the P1-change and performance was not significant ($r = 0.12$, $p = 0.65$).

DISCUSSION

Using EEG, we measured auditory brainstem and cortical activity during an implicit learning paradigm. Similar to previous studies of rapid learning (Abla et al., 2008; Alain et al., 2010; Ben-David et al., 2011), including a recent FFR study (Elmer et al., 2017), learning was associated with reduced neural activity. Although learning-related reduction occurred to some degree for both the easier and harder conditions, the two stimulus conditions showed different learning-related effects with respect to the brainstem and cortical components. In the easier condition, the FFR as well as the P1 and P2 components, differed from baseline, suggesting learning-related changes to both brainstem and cortex. The magnitude of the effect is visually similar for the two cortical components in the easier condition. Yet for the harder condition, only P1 and not the FFR or P2, differed from baseline, and this effect was driven by the learners in the sample. Thus, while P1 emerged as a potential marker of learning in both conditions, the FFR and P2 emerged only in the easier condition where learning was robust across participants. Our findings converge with other work to suggest that predictive coding occurs both cortically and subcortically but that it varies in its representation across brainstem and different cortical regions (Nieto-Diego and Malmierca, 2016; Font-Alaminos et al., 2021), and that it reflects long-term experience with sound that involves complex computations that go beyond low-level stimulus characteristics like inter-tone TPs (Kraus, 2021). Our constellation of findings also paints a complex picture of the possible timeline and top-down directionality of auditory network changes during rapid learning and they support the idea that learning emerge from multi-level representations of stimulus coding (Carbajal and Malmierca, 2018).

Cortical and brainstem potentials are both known to be sensitive to top-down effects, such as visual processing load (Xie et al., 2018), however, the involvement of the corticofugal pathway in these top-down effects and the learning-related processes studied here, remains inferential due to the lack of anatomic precision of the EEG signal. Techniques with greater anatomic and temporal precision do exist for studying human medial olivocochlear efferents, the lowest branch of auditory efferents. Human studies of medial olivocochlear function suggest that learning is dependent on efferent activity (de Boer and Thornton, 2008) and that the time scale of top-down modulation is fast (milliseconds to seconds) (Zhao and Dhar, 2011). Collectively, this supports the idea that top-down processes could potentially guide auditory learning over the brief timeframe of our rapid implicit learning paradigm (15-min of total exposure to the stimulus).

The important role of top-down processes in learning has been formalized in several popular models of learning and perception. The Reverse Hierarchy Theory of perceptual learning (Nahum et al., 2008; Ahissar et al., 2009), for example, proposes

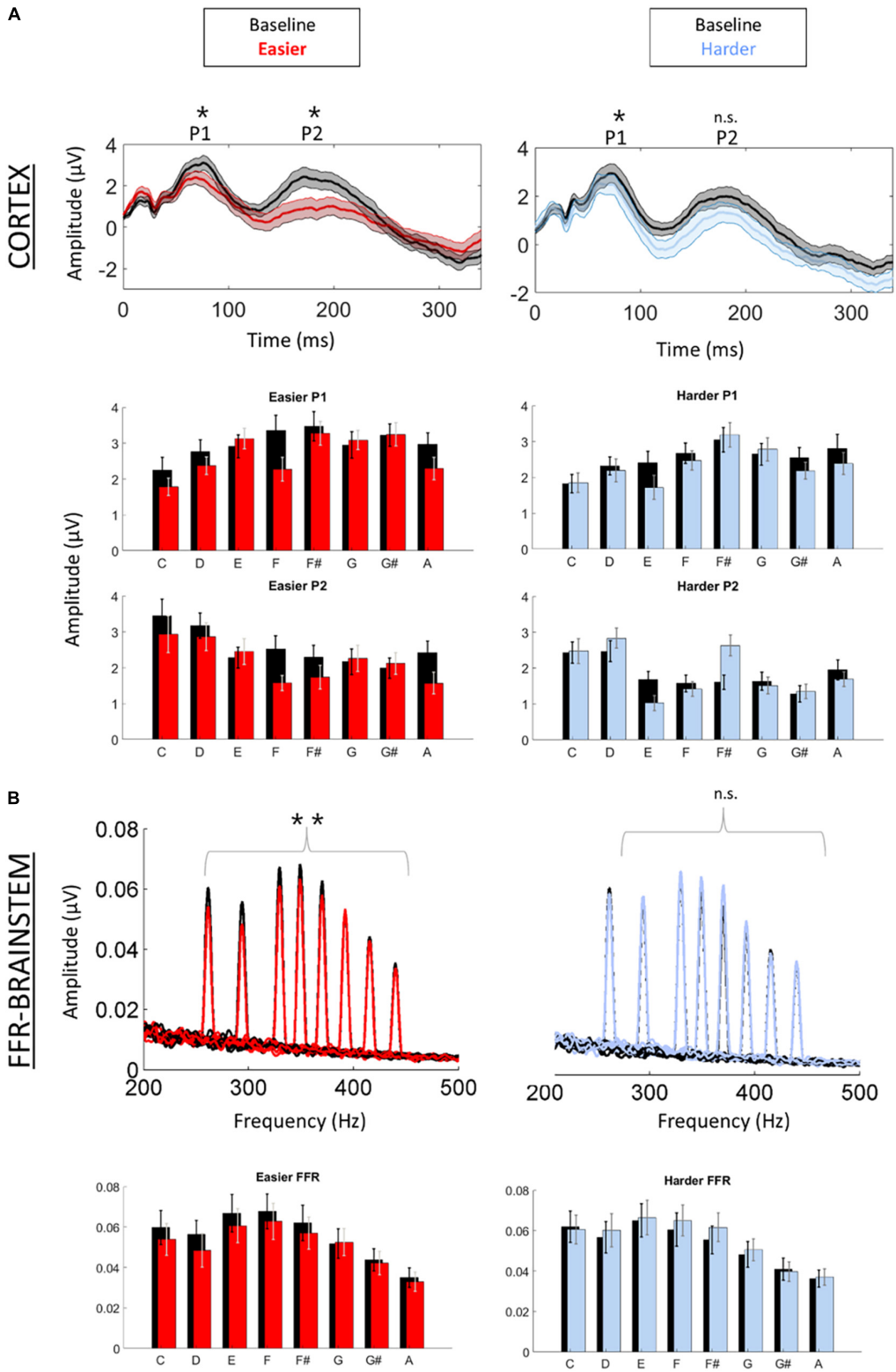

FIGURE 2 | Cortical (A) and brainstem (B) responses were recorded simultaneously while participants listened to a baseline (unpatterned) condition (black) followed by a patterned condition (red = easier, blue = harder). **(A)** Time-domain cortical evoked potentials are plotted for the baseline and patterned conditions (averaged across all tones). Shading represents + /1 standard error of the mean (SEM) amplitude. Peaks P1 and P2 are labeled. Bar graphs of the mean amplitude for each tone for P1 and P2 are plotted below the time domain waveforms. Error bars represent ± 1 SEM. **(B)** Brainstem frequency-following responses are plotted for each of the eight tones across both conditions. Bar graphs of the mean amplitude of the FFR for each tone are plotted below waveforms. Error bars represent ± 1 SEM. *$p < 0.05$; **$p < 0.01$.

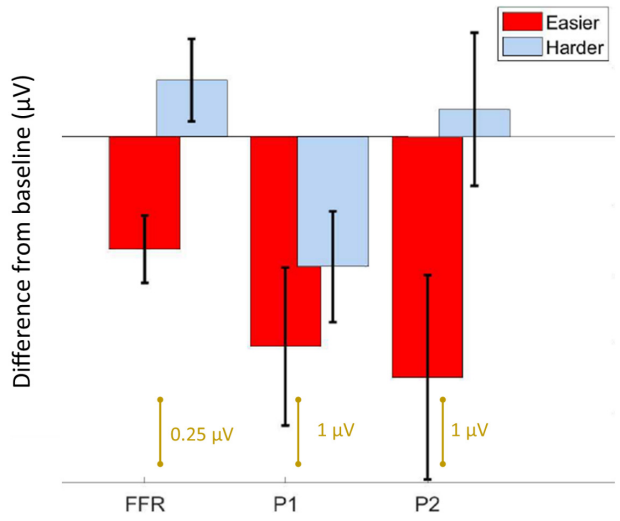

FIGURE 3 | Summary of results. Bar graph showing the average difference from baseline ± 1 standard error of the mean for the FFR, P1, and P2 components. For the easier condition, both the brainstem and cortical components are reduced from baseline, whereas for the harder condition, only P1 is different. Note the FFR is smaller in amplitude than the cortical components, and is plotted on a different scale.

that perception is controlled by top-down, sequential processes whereby the perceiver first detects the gestalt structure and then only later, at more expert stages of learning, is aware of the compositional elements of the structure. The Reverse Hierarchy Theory further postulates that lower-level sensory changes arise only at more advanced learning stages and that through a "backward cascade" fine-grained sensory detail can be retrieved from lower neuroanatomical structures. This conceptualization of auditory learning as a backward propagation from higher to lower structures throughout learning is consistent with a recent study showing that sensory changes, measured *via* the FFR, did not emerge until learners were overtrained on the task (Reetzke et al., 2018). It is also consistent with evidence from motor learning, where early learning is mediated first cortically and later subcortically (Penhune and Doyon, 2002). If our results are interpreted under the Reverse Hierarchy Theory framework, P1 could be viewed as the antecedent in a chain of learning-related events, with changes to the FFR and P2 reflecting a later stage(s) of learning. If so, longer exposure to the harder condition might allow learning to progress to a later stage, with brainstem (FFR) changes eventually emerging.

The top-down sequential view of learning that we adopt is also in alignment with the Propagation Hypothesis, which posits that the memory trace for a sound is propagated to earlier processing stages each time the stimulus is presented (Baldeweg, 2006). This sequential framing of physiological changes also echoes results from the animal literature where subcortical and cortical experience-dependent plasticity occurred on different timescales (Lu et al., 2014), with cortical changes being antecedent to subcortical changes. While speculative, our findings could indicate that subcortical changes are subordinate to cortical changes and arise through top-down, cortically guided predictive coding processes during auditory learning. However, we offer this conclusion with some caution given that the timeline of changes was tested only indirectly through a stimulus comparison and not through a longitudinal design. We also note that the need to average across trials hindered our ability to directly study the timeline of changes within the existing dataset. It is also important to acknowledge that predictive coding in the brainstem diminishes but does not vanish when the cortex is de-activated (Anderson and Malmierca, 2013), which conflicts with a fully top-down account of auditory learning.

An alternative explanation for our findings is that cortical neurons are inherently more plastic and more sensitive to stimulus regularities, and therefore change before the brainstem, without brainstem changes necessarily being cortically guided. A recent study, however, complicates the conclusion that brainstem is necessarily slower to change than cortex (Elmer et al., 2017). In that study, participants underwent 1-h of phonetic discrimination training, with FFRs measured before and after training. Discrimination improved with training relative to a passive listening group, with greater improvement correlating with larger FFR suppression. However, in contrast to the FFR, no changes were observed to the mismatch negativity response (MMN), a cortically-generated response linked to stimulus specific adaptation in response to repeated stimulation (Nieto-Diego and Malmierca, 2016). Unlike our study which recorded FFRs and cortical potentials simultaneously during the learning phase, EEG wasn't recorded during training, and the FFR and MMN were instead recorded sequentially using different stimulus paradigms. For the MMN, this involved an oddball paradigm using the end point stimuli of a phonetic continuum that produced robust MMNs even before training, suggesting possible ceiling effects for training on the MMN. The stimulus continuum was drawn from the participant's native language (German), so previous familiarity with the stimuli could have influenced the physiologic results. By contrast the tonal sequences in our experiment were entirely novel. However, one of them sounded more musical, likely because the four tone pairs created musical motifs (combinations of sounds) that are common in Western music. This increased musicality, we speculate, may have facilitated predictive coding and sped up the learning process, allowing for more learning to take place within the same period (∼15 min) (Skoe et al., 2015; Carbajal and Malmierca, 2018). Despite these differences and limitations, the Elmer et al. (2017) study converges with our work to suggest that brainstem and cortex are differentially sensitive to short-term physiological changes related to auditory learning.

Our results provide the foundation for futures studies into the time course and directionality of changes within the auditory neural network during implicit learning. Although this line of research is admittedly still in elementary stages, and results should be confirmed in longitudinal designs, the preliminary data we present here reinforce that no single brain region provides a comprehensive chronicle of what is involved in auditory learning, that stimulus statistics are not redundantly represented across the auditory system, and that auditory learning proceeds in stages, with subcortical changes emerging later.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Northwestern University Institutional Board. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

ESk designed the study, with input from all authors, and wrote the manuscript, with input and feedback from all authors. ESk and ESp collected the data. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported by the Knowles Hearing Center at Northwestern University.

ACKNOWLEDGMENTS

We would like to thank Beverly Wright for her discussion of the data and Scott Miller for sharing his stimuli. We would also like to thank Trent Nicol, Travis White-Schwoch, Karen Chan Barrett, and Samira Anderson for their comments on an earlier version of the manuscript.

REFERENCES

Abla, D., Katahira, K., and Okanoya, K. (2008). On-line assessment of statistical learning by event-related potentials. *J. Cogn. Neurosci.* 20, 952–964. doi: 10.1162/jocn.2008.20058

Ahissar, M., Nahum, M., Nelken, I., and Hochstein, S. (2009). Reverse hierarchies and sensory learning. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 364, 285–299. doi: 10.1098/rstb.2008.0253

Alain, C., Campeanu, S., and Tremblay, K. (2010). Changes in sensory evoked responses coincide with rapid improvement in speech identification performance. *J. Cogn. Neurosci.* 22, 392–403. doi: 10.1162/jocn.2009.21279

Anderson, L., and Malmierca, M. (2013). The effect of auditory cortex deactivation on stimulus-specific adaptation in the inferior colliculus of the rat. *Eur. J. Neurosci.* 37, 52–62. doi: 10.1111/ejn.12018

Bajo, V. M., Nodal, F. R., Moore, D. R., and King, A. J. (2010). The descending corticocollicular pathway mediates learning-induced auditory plasticity. *Nat. Neurosci.* 13, 253–260. doi: 10.1038/nn.2466

Baldeweg, T. (2006). Repetition effects to sounds: evidence for predictive coding in the auditory system. *Trends Cogn. Sci.* 10, 93–94. doi: 10.1016/j.tics.2006.01.010

Ben-David, B. M., Campeanu, S., Tremblay, K. L., and Alain, C. (2011). Auditory evoked potentials dissociate rapid perceptual learning from task repetition without learning. *Psychophysiology* 48, 797–807. doi: 10.1111/j.1469-8986.2010.01139.x

Carbajal, G. V., and Malmierca, M. S. (2018). The neuronal basis of predictive coding along the auditory pathway: from the subcortical roots to cortical deviance detection. *Trends Hear.* 22:2331216518784822.

Coffey, E. B., Nicol, T., White-Schwoch, T., Chandrasekaran, B., Krizman, J., Skoe, E., et al. (2019). Evolving perspectives on the sources of the frequency-following response. *Nat. Commun.* 10, 1–10. doi: 10.1007/978-3-319-47944-6_1

Crowley, K. E., and Colrain, I. M. (2004). A review of the evidence for P2 being an independent component process: age, sleep and modality. *Clin. Neurophysiol.* 115, 732–744. doi: 10.1016/j.clinph.2003.11.021

de Boer, J., and Thornton, A. R. (2008). Neural correlates of perceptual learning in the auditory brainstem: efferent activity predicts and reflects improvement at a speech-in-noise discrimination task. *J. Neurosci.* 28, 4929–4937. doi: 10.1523/jneurosci.0902-08.2008

Elmer, S., Hausheer, M., Albrecht, J., and Kühnis, J. (2017). Human brainstem exhibits higher sensitivity and specificity than auditory-related cortex to short-term phonetic discrimination learning. *Sci. Rep.* 7, 1–13.

Font-Alaminos, M., Ribas-Prats, T., Gorina-Careta, N., and Escera, C. (2021). Emergence of prediction error along the human auditory hierarchy. *Hear. Res.* 399:107954. doi: 10.1016/j.heares.2020.107954

Hornickel, J., Zecker, S. G., Bradlow, A. R., and Kraus, N. (2012). Assistive listening devices drive neuroplasticity in children with dyslexia. *Proc. Natl. Acad. Sci. U. S. A.* 109, 16731–16736. doi: 10.1073/pnas.1206628109

Jeng, F.-C., Costilow, C. E., Stangherlin, D. P., and Lin, C.-D. (2011). Relative Power of Harmonics in Human Frequency-Following Responses Associated with Voice Pitch in American and Chinese Adults. *Percept. Mot. Skills* 113, 67–86. doi: 10.2466/10.24.pms.113.4.67-86

Kraus, N. (2021). Memory for sound: the BEAMS Hypothesis. *Hear. Res.* 407:108291. doi: 10.1016/j.heares.2021.108291

Lu, H., Syka, J., Chiu, T.-W., and Poon, P. W. (2014). Prolonged sound exposure has different effects on increasing neuronal size in the auditory cortex and brainstem. *Hear. Res.* 314, 42–50. doi: 10.1016/j.heares.2014.05.005

Martin, B. A., Tremblay, K. L., and Korczak, P. (2008). Speech evoked potentials: from the laboratory to the clinic. *Ear Hear.* 29, 285–313. doi: 10.1097/aud.0b013e3181662c0e

Misyak, J. B., and Christiansen, M. H. (2012). Statistical Learning and Language: an Individual Differences Study. *Lang. Learn.* 62, 302–331. doi: 10.1111/j.1467-9922.2010.00626.x

Nahum, M., Nelken, I., and Ahissar, M. (2008). Low-level information and high-level perception: the case of speech in noise. *PLoS Biol.* 6:e126. doi: 10.1371/journal.pbio.0060126

Nieto-Diego, J., and Malmierca, M. S. (2016). Topographic distribution of stimulus-specific adaptation across auditory cortical fields in the anesthetized rat. *PLoS Biol.* 14:e1002397. doi: 10.1371/journal.pbio.1002397

Penhune, V. B., and Doyon, J. (2002). Dynamic cortical and subcortical networks in learning and delayed recall of timed motor sequences. *J. Neurosci.* 22, 1397–1406. doi: 10.1523/jneurosci.22-04-01397.2002

Picton, T. W., and Hillyard, S. A. (1974). Human auditory evoked potentials. II. Effects of attention. *Electroencephalogr. Clin. Neurophysiol.* 36, 191–199. doi: 10.1016/0013-4694(74)90156-4

Ponton, C., Eggermont, J. J., Khosla, D., Kwong, B., and Don, M. (2002). Maturation of human central auditory system activity: separating auditory evoked potentials by dipole source modeling. *Clin. Neurophysiol.* 113, 407–420. doi: 10.1016/s1388-2457(01)00733-7

Reetzke, R., Xie, Z., Llanos, F., and Chandrasekaran, B. (2018). Tracing the trajectory of sensory plasticity across different stages of speech learning in adulthood. *Curr. Biol.* 28, 1419–1427.e4.

Saffran, J. R. (2003). Musical learning and language development. *Ann. N. Y. Acad. Sci.* 999, 397–401. doi: 10.1196/annals.1284.050

Saffran, J. R., Newport, E. L., Aslin, R. N., Tunick, R. A., and Barrueco, S. (1997). Incidental Language Learning: listening (And Learning) out of the Corner of Your Ear. *Psychol. Sci.* 8, 101–105. doi: 10.1111/j.1467-9280.1997.tb00690.x

Saffran, J. R., Johnson, E. K., Aslin, R. N., and Newport, E. L., et al. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition* 70, 27–52. doi: 10.1016/S0010-0277(98)00075-4

Skoe, E., and Kraus, N. (2010). Auditory brain stem response to complex sounds: a tutorial. *Ear Hear.* 31, 302–324. doi: 10.1097/aud.0b013e3181cdb272

Skoe, E., Krizman, J., Spitzer, E., and Kraus, N. (2015). Prior experience biases subcortical sensitivity to sound patterns. *J. Cogn. Neurosci.* 27, 124–140. doi: 10.1162/jocn_a_00691

Song, J. H., Skoe, E., Banai, K., and Kraus, N. (2012). Training to improve hearing speech in noise: biological mechanisms. *Cereb. Cortex* 22, 1180–1190. doi: 10.1093/cercor/bhr196

Suga, N., Xiao, Z., Ma, X., and Ji, W. (2002). Plasticity and corticofugal modulation for hearing in adult animals. *Neuron* 36, 9–18. doi: 10.1016/s0896-6273(02)00933-9

Sussman, E., Steinschneider, M., Gumenyuk, V., Grushko, J., and Lawson, K. (2008). The maturation of human evoked brain potentials to sounds presented at different stimulus rates. *Hear. Res.* 236, 61–79. doi: 10.1016/j.heares.2007.12.001

Tichko, P., and Skoe, E. (2017). Frequency-dependent fine structure in the frequency-following response: the byproduct of multiple generators. *Hear. Res.* 348, 1–15. doi: 10.1016/j.heares.2017.01.014

Toro, J. M., and Trobalon, J. B. (2005). Statistical computations over a speech stream in a rodent. *Percept Psychophys* 67, 867–875. doi: 10.3758/BF03193539

Tremblay, K., Kraus, N., McGee, T., Ponton, C., and Otis, B. (2001). Central auditory plasticity: changes in the N1-P2 complex after speech-sound training. *Ear Hear.* 22, 79–90. doi: 10.1097/00003446-200104000-00001

Tremblay, K. L., and Kraus, N. (2002). Auditory training induces asymmetrical changes in cortical neural activity. *J. Speech Lang. Hear. Res.* 45, 564–572. doi: 10.1044/1092-4388(2002/045)

Weiss, D. J., Gerfen, C., and Mitchel, A. D. (2009). Speech Segmentation in a Simulated Bilingual Environment: a Challenge for Statistical Learning? *Lang. Learn. Dev.* 5, 30–49. doi: 10.1080/15475440802340101

White-Schwoch, T., Krizman, J., Nicol, T., and Kraus, N. (2021). Case studies in neuroscience: cortical contributions to the frequency-following response depend on subcortical synchrony. *J. Neurophysiol.* 125, 273–281. doi: 10.1152/jn.00104.2020

Winer, J. A. (2006). Decoding the auditory corticofugal systems. *Hear. Res.* 212, 1–8. doi: 10.1016/j.heares.2005.06.014

Xie, Z., Reetzke, R., and Chandrasekaran, B. (2018). Taking attention away from the auditory modality: context-dependent effects on early sensory encoding of speech. *Neuroscience* 384, 64–75. doi: 10.1016/j.neuroscience.2018.05.023

Zhao, W., and Dhar, S. (2011). Fast and slow effects of medial olivocochlear efferent activity in humans. *PLoS One* 6:e18725. doi: 10.1371/journal.pone.0018725

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Skoe, Krizman, Spitzer and Kraus. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Neuroscience

REVIEW
published: 19 August 2021
doi: 10.3389/fnins.2021.723893

Lemniscal Corticothalamic Feedback in Auditory Scene Analysis

Natsumi Y. Homma[1,2]* *and Victoria M. Bajo*[3]

[1] Center for Integrative Neuroscience, University of California, San Francisco, San Francisco, CA, United States, [2] Coleman Memorial Laboratory, Department of Otolaryngology – Head and Neck Surgery, University of California, San Francisco, San Francisco, CA, United States, [3] Department of Physiology, Anatomy and Genetics, University of Oxford, Oxford, United Kingdom

Sound information is transmitted from the ear to central auditory stations of the brain via several nuclei. In addition to these ascending pathways there exist descending projections that can influence the information processing at each of these nuclei. A major descending pathway in the auditory system is the feedback projection from layer VI of the primary auditory cortex (A1) to the ventral division of medial geniculate body (MGBv) in the thalamus. The corticothalamic axons have small glutamatergic terminals that can modulate thalamic processing and thalamocortical information transmission. Corticothalamic neurons also provide input to GABAergic neurons of the thalamic reticular nucleus (TRN) that receives collaterals from the ascending thalamic axons. The balance of corticothalamic and TRN inputs has been shown to refine frequency tuning, firing patterns, and gating of MGBv neurons. Therefore, the thalamus is not merely a relay stage in the chain of auditory nuclei but does participate in complex aspects of sound processing that include top-down modulations. In this review, we aim (i) to examine how lemniscal corticothalamic feedback modulates responses in MGBv neurons, and (ii) to explore how the feedback contributes to auditory scene analysis, particularly on frequency and harmonic perception. Finally, we will discuss potential implications of the role of corticothalamic feedback in music and speech perception, where precise spectral and temporal processing is essential.

Keywords: descending projections, layer VI cortical neurons, thalamus, medial geniculate body, tonotopy, harmonicity, speech, music

OPEN ACCESS

Edited by:
David K. Ryugo,
Garvan Institute of Medical Research, Australia

Reviewed by:
Jennifer F. Linden,
University College London, United Kingdom
Kexin Yuan,
Tsinghua University, China

***Correspondence:**
Natsumi Y. Homma
Natsumi.Homma@ucsf.edu

Specialty section:
This article was submitted to Auditory Cognitive Neuroscience, a section of the journal Frontiers in Neuroscience

Received: *11 June 2021*
Accepted: *30 July 2021*
Published: *19 August 2021*

Citation:
Homma NY and Bajo VM (2021) Lemniscal Corticothalamic Feedback in Auditory Scene Analysis. Front. Neurosci. 15:723893. doi: 10.3389/fnins.2021.723893

INTRODUCTION

In everyday life we are constantly analyzing our acoustic environment, which is filled with sounds from many different sources. For example, we can listen to a person next to us while others in the room are chatting. With little effort, we can treat the voice of that talker as the desired foreground signal and segregate it from the background of all other sounds in the room. We can also listen to the melody in a symphony while focusing on the parts played by different musical instruments. These listening abilities are based on the process of "auditory scene analysis" (Bregman, 1990), which is essential for grouping or segregating sound mixtures into perceptually meaningful categories while perceiving the whole auditory environment.

In auditory scene analysis, we categorize different, simultaneously occurring sounds based on several acoustic properties such as onset asynchrony (Darwin, 1984; Darwin and Ciocca, 1992),

harmonicity (Moore et al., 1985; Hartmann et al., 1990), and spatial information (Cherry, 1953; Eramudugolla et al., 2008). We can also group sound sequences, for example, based on their rhythmic structure (Miller and Heise, 1950; Bregman and Campbell, 1971; van Noorden, 1975). In addition to exploiting various acoustic attributes, scene analysis relies on attention, learning and memory and incorporates information from other sensory systems (Shamma et al., 2011; Maddox et al., 2015; Atilgan et al., 2018).

Human speech, animal vocalizations, and the sounds produced by many musical instruments are all periodic sounds comprised of a fundamental frequency (F0) plus its multiple integer harmonics (**Figure 1**). This condition is referred as "harmonicity" and is one of the basic acoustic properties for auditory scene analysis. In addition, harmonicity is considered as a basis of the perceptual attribute, "pitch," that allow us to order sounds from low to high. Intonation, rising or falling of pitch, expresses grammatical meaning or emotion in speech. Pitch also helps in discriminating voices of different speakers, identifying different musical instruments, or conveying the melodic line in music. F0 differences between concurrent vowels contribute to indicate and segregate different talkers (Culling and Darwin, 1993; de Cheveigné, 1995; Arehart et al., 2011). The harmonic components of vowels show a unique energy distribution of frequencies, known as the spectral envelope. The peaks in the envelope ("formant frequencies") characterize phonemes and provide critical spectral information to discriminate them (Klatt, 1982; Swanepoel et al., 2012). Similarly, the spectral envelope is characteristic for an individual musical instrument and defines its sound quality ("timbre") (**Figure 1D**; Town and Bizley, 2013). Harmonicity and spectral regularity can serve as a strong grouping cue and play a key role for music and speech perception (McDermott and Oxenham, 2008; Micheyl and Oxenham, 2010). The harmonic structure of the sound spectrum helps the listener keep track of a speaker in competing, simultaneous speech signals (de Cheveigné et al., 1997; Popham et al., 2018). Moreover, harmonicity can contribute to longer lasting memory storage by connecting several aspects of spectral information to the behavior of a single attribute, namely F0 (McPherson and McDermott, 2020). Although spectral regularity indicates harmonic relations, it can be extended to simply regular intervals of spectral components that deviate from multiple integer harmonics of F0 but have equal spacing or that include shifting of phases on each component, while holding grouping effect and behavioral relevance (Roberts and Bailey, 1996; So et al., 2020).

Natural sounds, such as human speech, animal vocalizations, and many environmental sounds, are comprised of complex spectrotemporal modulations (Chi et al., 1999; Elliott and Theunissen, 2009; McDermott and Simoncelli, 2011). Modulation energy in speech, for example, is captured by a decaying low-pass distribution below ∼16 Hz temporal modulation frequencies and ∼2 cycle/octave spectral modulation frequencies (Chi et al., 1999; Elliott and Theunissen, 2009). Word recognition is severely impaired when temporal modulation frequencies below ∼8–12 Hz are unavailable (Drullman et al., 1994; Elliott and Theunissen, 2009). This upper limit closely corresponds to the syllabic rate in speech (Greenberg et al., 2003; Pellegrino et al., 2011). Identifying syllables is essential for distinguishing words. The analysis of Western music revealed that temporal modulation for music is similar to speech; however, the peak energy is shifted down from 5 Hz for speech to 3 Hz for music reflecting the typical tempo (beats or rhythms) of music (Ding et al., 2017). Both for speech and music, most of modulation energy is <32 Hz, although rapid temporal modulations >50 Hz are critical for the perception of aspects such as pitch, lexical meaning, formant and timbre patterns, and their detections in noisy environments (Rosen, 1992; Shamma and Lorenzi, 2013).

Sound information is transmitted from the cochlea, via the medulla, pons, midbrain, and thalamus to the cortex. In parallel to the ascending pathways, descending pathways project information back to each stage (Winer, 2006; Souffi et al., 2021). Feedback projection activity can alter neural excitability (Villa et al., 1991; He, 1997, 2003; Lohse et al., 2020) and tuning properties for frequency (Yan and Ehret, 2002; Luo et al., 2011), intensity (Yan and Ehret, 2002; Ma and Suga, 2007), sound onset (Luo et al., 2008), sound duration (Ma and Suga, 2007), or sound source location (Nakamoto et al., 2008) as has been shown for the cochlear nuclei, inferior colliculus (IC), or medial geniculate body (MGB). Behavioral experiments have demonstrated that descending feedback projections modulate perceptual abilities, including detection and discrimination of sound frequency (Guo et al., 2017), harmonicity (Homma et al., 2017), and location (Bajo et al., 2010) (for review; Lohse et al., 2019).

The projection from layer VI neurons of primary auditory cortex (A1) to the ventral division of medial geniculate body (MGBv) is one of the major feedback pathways (**Figure 2A**). A1-MGBv corticothalamic feedback projections have small excitatory terminals (Ojima, 1994), which are thought to act as "modulators" that regulate gain and firing patterns in the thalamus (Sherman and Guillery, 1998, 2011). The balance of excitation from the corticothalamic neurons and inhibition from thalamic reticular nucleus (TRN) neurons, which receives collateral projections from thalamocortical and corticothalamic neurons, modulates the tuning properties, gating and firing patterns of MGBv neurons (Zhang and Suga, 2000; Tang et al., 2012; Guo et al., 2017; Lohse et al., 2020).

In this review, we first introduce the anatomy and physiology of A1-MGBv corticothalamic feedback projections, then we summarize the feedback modulations of thalamic responses related to spectral and temporal information for speech and music processing. Finally, we explore perceptual effects of corticothalamic feedback, particularly on frequency and harmonicity analysis, and discuss how A1-MGBv corticothalamic feedback could contribute to our music and speech perception.

A1-MGBv CORTICOTHALAMIC PATHWAY

Ascending Thalamocortical Projections

The main auditory nucleus of the thalamus, i.e., the MGB, is subdivided into three distinct areas. The ventral division is part of the lemniscal pathway and shows a tonotopic organization,

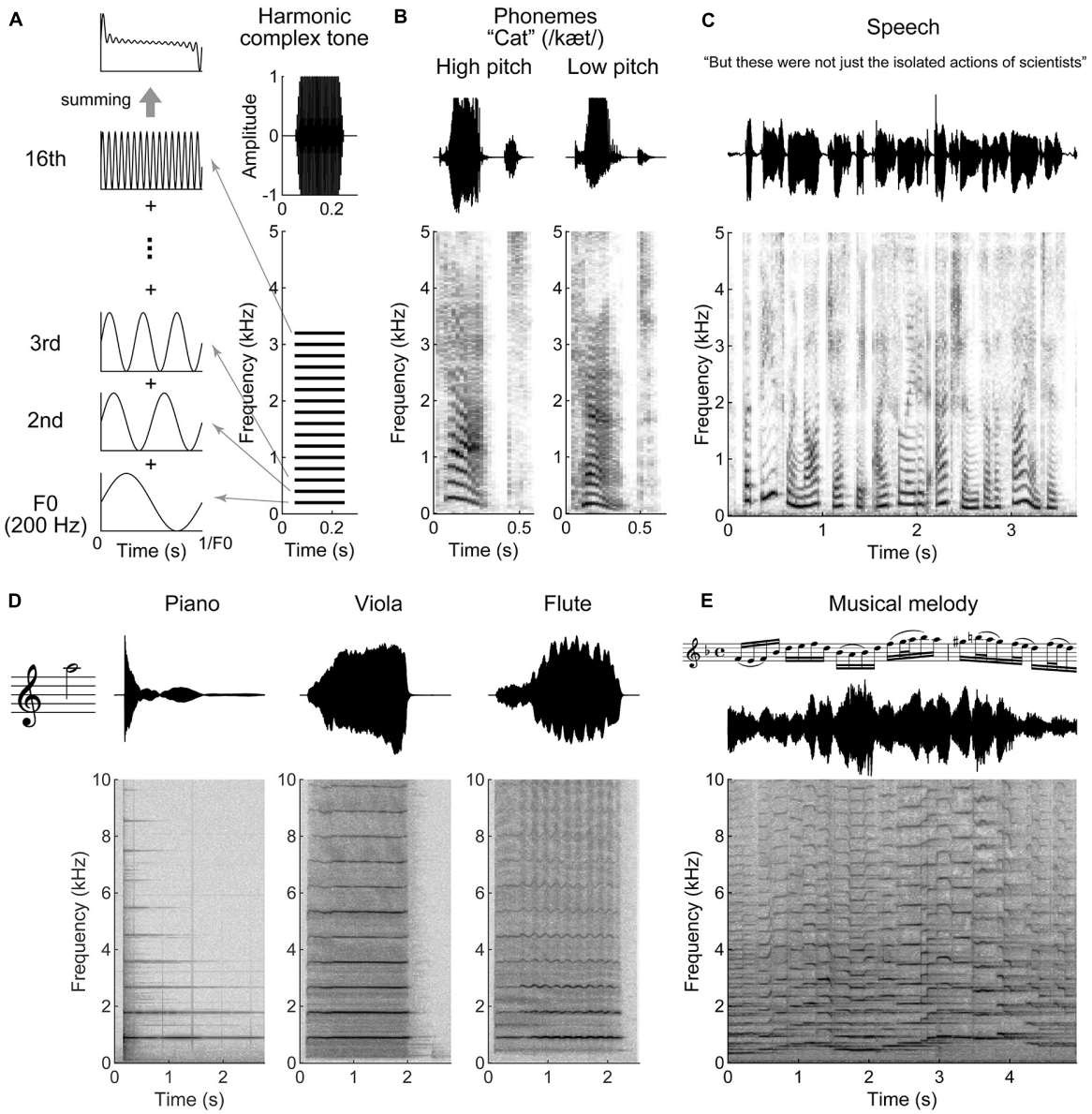

FIGURE 1 | Harmonic structures in speech and music. **(A)** Schematic illustration of a harmonic complex tone comprised 200 Hz fundamental frequency (F0) with 16 harmonics (right bottom). The 16 harmonics are integer multiples of F0. The right top indicates a total amplitude waveform of the tone. The left panels show waveforms of F0, 2nd, 3rd, and 16th harmonics. The periodicity of the harmonic complex tone, the sum of all the harmonics (left top panel), is equal to F0. **(B–E)** Waveforms (top) and spectrograms (bottom) for a segment of speech and music. **(B)** Female (high pitch) and male (low pitch) voices pronouncing, "cat." The harmonic structures are observed around the vowel "a" (æ). **(C)** A segment of speech. Syllables occur every ∼0.2 to 0.5 s (2–5 Hz). **(D)** The instrumental tones played by piano, viola, and flute at 880 Hz F0 (A5 note). Although the tones evoke the same pitch sensation, their timbre differs as indicated with the waveforms and spectrograms. **(E)** A segment of musical melody played by violin. The tone pitch fluctuates faster than speech in this example (it could be slower dependent on tempo and rhythm).

in which characteristic frequencies of neurons are arranged in a dorsolateral to ventromedial topographic gradient from low to high frequencies (Aitkin and Webster, 1972; Calford, 1983; **Figure 2A**). MGBv mainly receives inputs from the ipsilateral central nucleus of the inferior colliculus (CNIC) (Calford and Aitkin, 1983; LeDoux et al., 1985; Rouiller and de Ribaupierre, 1985) with glutamatergic but also GABAergic connections (Winer et al., 1996; Peruzzi et al., 1997) and projects to A1 (Andersen et al., 1980a; Lee and Winer, 2008). The cerebral cortex consists of different types of cells that are functionally organized into a laminar structure, and MGBv neurons mainly target layer III/IV of A1 but also other cortical layers and especially layer I (Huang and Winer, 2000; Kimura et al., 2003; Smith et al., 2012; Vasquez-Lopez et al., 2017; Xie et al., 2018). MGBv also receives a small amount of input from the shell regions of IC (Kudo and Niimi, 1980; LeDoux et al., 1985) and

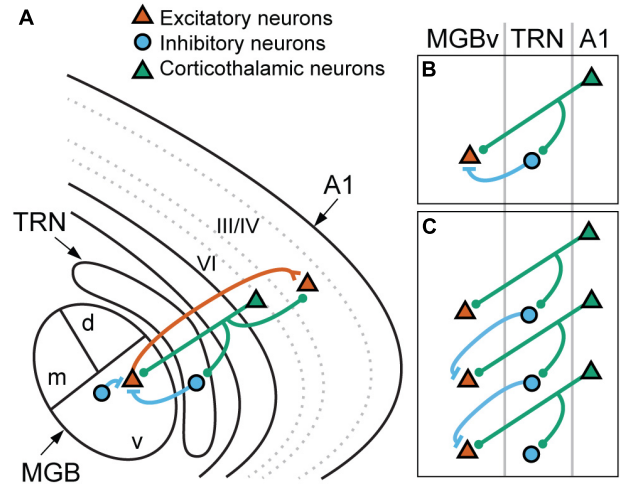

FIGURE 2 | Corticothalamic connections to the lemniscal thalamus mediated via the thalamic reticular nucleus. **(A)** Schematic illustration of thalamocortical and corticothalamic connections. Thalamocortical neurons (red) in the ventral division of medial geniculate body (MGBv) project mainly to the neurons in layer III/IV of the primary auditory cortex (A1). The collaterals innervate to the thalamic reticular nucleus (TRN) but are not depicted in the diagram for simplicity. Corticothalamic neurons (green) in layer VI project back to the MGBv with the collaterals to the TRN and layer III/IV of A1. TRN and local GABAergic neurons provide inhibitory inputs (blue) to MGBv neurons. **(B,C)** Potential roles of the corticothalamic feedback. **(B)** Corticothalamic neurons modulate gain of MGBv neurons by regulating the balance of monosynaptic excitatory and disynaptic inhibitory inputs. **(C)** Corticothalamic neurons modulate tuning properties of MGBv neurons by lateral inhibition. The collaterals onto TRN spread along neighboring neurons with similar but not identical BFs, and the excitatory and inhibitory inputs to the MGBv neurons come from different corticothalamic cells. d, dorsal MGB; m, medial MGB; v, ventral MGB.

innervates other "core" cortical areas that usually express a tonotopic organization and, to a lesser extent, "belt" areas that receive major non-lemniscal thalamic inputs (Andersen et al., 1980a; Morel and Imig, 1987; Redies et al., 1989; Velenovsky et al., 2003; De La Mothe et al., 2006; Polley et al., 2007; Lee and Winer, 2008; Hackett et al., 2011; Saldeitis et al., 2014). Thalamic inputs sourcing different locations of MGBv in the caudal-to-rostral dimension project to the targeted locations of the core cortical areas in the ventral-to-dorsal dimension without much overlap (Storace et al., 2010, 2011; Read et al., 2011), suggesting distinct functional segregations within the lemniscal thalamocortical system.

By contrast, the medial and dorsal divisions of the auditory thalamus (MGBm and MGBd) are considered non-lemniscal areas and do not have a clear tonotopic organization compared to MGBv (Calford, 1983; Rouiller et al., 1989; Hackett et al., 2011). Comparative studies of these three divisions and non-lemniscal corticothalamic feedback have been reviewed previously (Bartlett, 2013; Lee, 2015). Briefly, MGBm mainly receives inputs from the external cortex of the IC and projects to all layers of the auditory cortex whereas MGBd is innervated by dorsal cortex of IC and projects to layers I, III/IV, VI of belt and parabelt regions and only weakly to core auditory areas (Andersen et al., 1980b; Kudo and Niimi, 1980; LeDoux et al., 1985). Non-lemniscal MGB also receives inputs from superior colliculus (Holstege and Collewijn, 1982), and sends outputs to the striatum and amygdala (LeDoux et al., 1991), integrating multimodal sensory and emotional information (Weinberger, 2011).

Descending Corticothalamic Projections

The tonotopic organization can also be a hallmark for descending pathways and is preserved in descending lemniscal axons. In principle, lemniscal cortical regions project back to subcortical lemniscal stations, whereas non-lemniscal cortical regions target non-lemniscal stations (Winer, 2006). Thus, lemniscal corticothalamic neurons project back from layer VI to MGBv with minor inputs to MGBm and MGBd (Andersen et al., 1980a; Rouiller and de Ribaupierre, 1985; Bajo et al., 1995; Budinger et al., 2013) in parallel to the tonotopic organization of lemniscal ascending thalamocortical projections (Redies et al., 1989; Rodrigues-Dagaeff et al., 1989; Velenovsky et al., 2003; Read et al., 2008; Hackett et al., 2011; Storace et al., 2011). The information flow is organized in a layer specific manner; layer VI corticothalamic neurons send collaterals to layer III/IV (**Figure 2A**; Ojima, 1994; Llano and Sherman, 2008), where the axon terminals of thalamocortical projections are mainly found. The A1-MGBv corticothalamic projections are mainly ipsilateral and form the focal topographic reciprocal connections; however, a minority showed non-reciprocal inputs (Winer and Larue, 1987; Winer et al., 2001). In addition, A1 also projects to MGBd as a descending feedforward projection, which is thought to transmit sound information from layer V of A1 to non-primary higher cortices (Ojima, 1994; Llano and Sherman, 2008). This cortico-thalamo-cortical connection is essential for corticocortical communication and processing higher-order sound features (Lee, 2015; Williamson and Polley, 2019).

The corticothalamic feedback projections from layer VI and the cortico-thalamo-cortical signaling from layer V are distinguished by their morphological characteristics. Tracer injection studies showed that the first-order corticothalamic feedback neurons in layer VI have small distal terminals with thin axons and convergent endings in MGBv while the higher-order corticothalamic feedforward neurons in layer V have large boutons in MGBd (Rouiller and Welker, 1991; Ojima, 1994; Bajo et al., 1995; Bartlett et al., 2000; Winer and Prieto, 2001). These two thalamic terminal types are characteristic for Class1 and Class2 glutamatergic projection neurons, respectively, based on structural and physiological properties (Sherman and Guillery, 1998, 2011). Corticothalamic feedforward neurons in layer V are classified as Class1, which have been characterized as "drivers" due to their function of relaying information to the cortex. On the other hand, corticothalamic feedback neurons in layer VI are classified as Class2 or "modulators" controlling how relay neurons transmit their information. They activate type I metabotropic glutamate receptors, which are identified as a characteristic of modulator synapses, and show paired-pulse facilitation with small excitatory post-synaptic potentials (EPSPs) in MGBv and layer IV cortical neurons (Bartlett and Smith, 2002; Lee and Sherman, 2009; Lee et al., 2012). A recent study also supports the idea that corticothalamic feedback

projection neurons in layer VI are "modulators" with high selectivity of information flow while corticothalamic feedforward projections from layer V are "drivers" that integrate and transmit information (Williamson and Polley, 2019). Additionally, layer VI corticothalamic neurons can be activated by preparatory motor actions that trigger reward and auditory inputs in behaving mice, supporting the modulatory role for active listening (Clayton et al., 2021). Overall, layer VI corticothalamic neurons modulate MGBv neurons by regulating their excitability, voltage gated conductance, and synaptic potentials.

Inhibitory Inputs and Intracortical Local Circuits

In addition to MGBv, corticothalamic feedback neurons send their axonal projection terminals to layer IV and the TRN (Jones, 2007; Sherman and Guillery, 2013; **Figure 2A**). TRN consists of GABAergic cells and is located between the cortex and the thalamus wrapping the thalamic nuclei with a sheet structure. The auditory sector is identified at the posterior-ventral part of TRN and receives thalamocortical collaterals (Rouiller et al., 1985; Villa, 1990; Conley et al., 1991). MGBv neurons receive inhibitory inputs from TRN in addition to local and IC GABAergic inputs (Winer and Larue, 1996; Winer et al., 1996; Arcelli et al., 1997; Peruzzi et al., 1997; Zhang et al., 2008; Clarke and Lee, 2018). The proportion of local GABAergic neurons in MGBv present species-specific variations, being almost absent in rodents and ∼25% of MGBv neurons in carnivores and primates (Winer and Larue, 1996). Corticothalamic projections to MGBv and TRN generally preserve topographic connections (Conley et al., 1991; Kimura et al., 2005; Cotillon-Williams et al., 2008); therefore, the feedback from layer VI shapes thalamic tuning by modulating the balance between converging excitation and inhibition.

In the cortex, the thalamocortical inputs are received mainly in layer III/IV. The thalamorecipient neurons then innervate the upper or supragranular cortical layers. Recent studies suggest that corticothalamic projection neurons in layer VI induce overall gain change across all cortical layers by recruiting local fast-spiking inhibitory neurons to modulate cortical oscillation (Olsen et al., 2012; Bortone et al., 2014; Kim et al., 2014; Guo et al., 2017). Taken together, this corticothalamic loop arborizing through A1, MGBv, and TRN shapes the receptive field structures and modulates information flow both in the thalamus and cortex and, thus, embodies a crucial aspect of the thalamocortical interface.

CORTICAL MANIPULATION ONTO MGBv NEURONS

Changes in Spectral Tuning

In laboratory experiments, single frequency tones ("pure tones") are generated as a sound with a sinusoidal waveform. A simple spectral tuning curve is typically obtained by measuring the change in response to each pure tone at different combinations of frequency and intensity. Neurons in MGBv predominantly show narrow tuning curves, indicating high frequency selectivity, and short response latencies (cats: Calford, 1983; Morel et al., 1987; Miller et al., 2002; rats: Bordi and LeDoux, 1994; guinea pigs: Edeline et al., 1999; mice: Anderson and Linden, 2011; marmosets: Bartlett et al., 2011), Overall, MGBv neurons show sharper frequency tuning compared to A1 neurons by 0.1–0.3 octave (Miller et al., 2002; Bartlett et al., 2011). It is unknown at this time whether the thalamic frequency tuning properties in anesthetized or passively listening animals are predominantly inherited from the midbrain or shaped by local circuits with or without corticothalamic feedback.

Thalamic tuning properties can be altered by manipulating the corticothalamic activities in layer VI. Modulation of frequency tuning by corticofugal activity is dependent on the relationship between the best frequencies (BFs), i.e., the frequency that evokes the highest firing rate, of subcortical and cortical neurons (e.g., Zhang and Suga, 2000; **Figure 3**). Observed tuning changes in MGBv neurons following A1 manipulation fall into two main categories. First, when the BF of a corticofugal neuron is matched to that of the recipient thalamic neuron, the frequency tuning at the latter is sharpened by facilitation of the responses at BF and reduction of responses to frequencies away from BF (**Figure 3A**). Conversely, when the BF of corticofugal neurons differs from that of the affected MGBv neuron, responses at the BF of MGBv neuron are reduced and responses to surrounding frequencies are enhanced, shifting the tuning curve away from the BF of the stimulated corticofugal neuron (**Figure 3B**). Cortical focal electric stimulation has been shown to induce both types of changes in MGBv of bats (Zhang and Suga, 2000; Tang et al., 2012). Similar changes are also induced in MGBv by electrical stimulation of the cholinergic nucleus basalis or by behavioral conditioning, and are abolished with inactivation of the auditory cortex (guinea pigs: Edeline and Weinberger, 1991; bats: Zhang et al., 1997; mice: Zhang and Yan, 2008; Luo et al., 2011; Nelson et al., 2015). This suggests that the effects on subcortical activity are controlled by the balance of local excitation and inhibition including the influence of various neuromodulators.

These two types of response changes potentially could arise from different, converging projections. When corticothalamic and corticoreticular pathways are strictly reciprocal, the balance of excitation and inhibition on a thalamic relay cell could be modulated by monosynaptic inputs from layer VI corticothalamic cells and disynaptic inputs from TRN mediated via the collateral of the same corticothalamic cells (**Figure 2B**). When the collaterals spread to the neighboring thalamic cells that have different tuning properties (**Figure 2C**), the corticoreticular pathway could sharpen thalamic receptive fields by lateral inhibition. Although the cascades of information flow from the periphery to the first-order thalamic nuclei differ among different sensory systems, the anatomy and physiology of thalamocortical and corticothalamic neurons are generally comparable (Rouiller and Welker, 2000; Sherman and Guillery, 2013). In the somatosensory system, most corticothalamic and corticoreticular projections are organized in a reciprocal manner and contribute to gain control on the thalamic cells, and some projections diverge to neighboring cells that have different tuning properties and show lateral inhibition (Temereanca and Simons, 2004; Li and Ebner, 2007; Lam and Sherman, 2010; Crandall et al.,

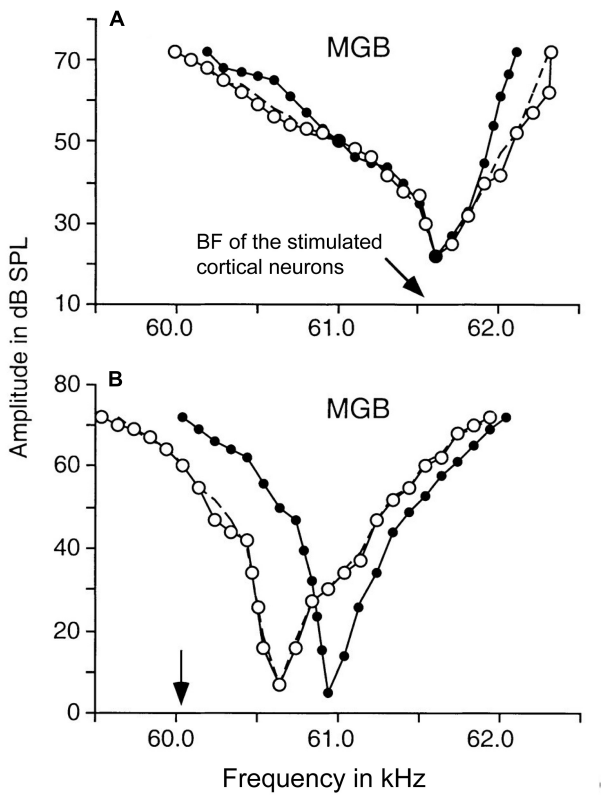

FIGURE 3 | Corticothalamic modulation of frequency tuning curves in MGBv neurons. A focal electrical activation of cortical neurons in the Doppler-shifted constant frequency processing area of Jamaican mustached bats evoked changes of frequency tuning curves in MGBv neurons. The black arrows indicate the best frequencies (BFs) of cortical neurons stimulated. The tuning curves were estimated before (white circles), during (black circles), and after (dashed lines) the cortical stimulation. **(A)** A MGBv neuron showed sharpening of the tuning curve when cortical neurons that had matched BF to the recipient MGBv neuron. BF of the MGBv neuron did not change. **(B)** Another MGBv neuron showed a reduction of responses around BF of the MGBv neuron when BF of stimulated cortical neurons is unmatched to the recipient MGBv neuron. The tuning curve was shifted away from the BF of stimulated cortical neurons with increasing responses outside of BF of the MGBv neuron. The changes were transient only lasting 1–2 h after the 7-min cortical stimulation (Adapted and modified with permission from Zhang and Suga, 2000; Figure 5).

2015). Orientation tuning of visual thalamic receptive fields also is shifted by focal pharmacological activation in layer VI of the primary visual cortex (Wang et al., 2016). Thus, a shift of tuning properties is a universal aspect of corticothalamic interactions and likely is mediated by disynaptic inhibition via diverged corticoreticular projections.

The ability to alter receptive field tuning may enable an attentive listener to focus on a specific speaker, other sound attributes of current interest, or adjust to changes in the sound environment by adapting to sound statistics in the environment similar to what has been demonstrated in cortical neurons (Fritz et al., 2003; Holdgraf et al., 2016; Homma et al., 2020). A major purpose of the modulation of receptive field selectivity by corticothalamic projections could be an increased discrimination ability, e.g., for frequency and musical pitch, to enhance scene analysis in complex sound environments.

Gain Changes

In addition to the modulation of tuning properties, it was observed that lemniscal corticothalamic feedback generally facilitates excitability of MGBv neurons (cats: Ryugo and Weinberger, 1976; Villa et al., 1991; He, 1997; guinea pigs: He et al., 2002; Yu et al., 2004; mice: Guo et al., 2017; Lohse et al., 2020; marmosets: Zhang et al., 2021). Deactivation of the auditory cortex decreases the spontaneous firing rate in MGBv neurons (Ryugo and Weinberger, 1976; Villa et al., 1991; but see Zhang et al., 2021) while cortical activation enhances their responses to sounds (He, 1997; He et al., 2002; Guo et al., 2017). The effect is, however, heterogeneous, and a minority of MGBv neurons can show suppression. The proportion of excitation and inhibition differs among species due to different density of local GABAergic interneurons in MGBv (Winer and Larue, 1996; Arcelli et al., 1997). Since local inhibitory cells as well as long-range inhibitory inputs from the TRN play a key role in shaping thalamic processing, the interspecies differences in inhibitory capacity may bear on potential differences in the ability to modify processing via corticothalamic inputs.

A recent study showed that activation of corticothalamic neurons in layer VI enhanced or suppressed activity in A1, MGBv, and TRN neurons with a dependence on relative timing of optogenetic and sound stimulation (Guo et al., 2017; **Figure 4**). The authors further found that the facilitated activities in MGBv could improve performance on a behavioral frequency detection task by increasing sound-evoked responses whereas better frequency discrimination was more closely related to suppressed cortical stimulus response (Olsen et al., 2012; Bortone et al., 2014). These observations indicate that corticothalamic feedback modulations are diverse and dependent on stimulus and/or task context.

The gain control provide by corticothalamic feedback induces a change in tuning sharpness by shifting the overall excitability of MGBv neurons but generally without affecting BF. As anatomical and physiological studies have shown (see section "A1-MGBv Corticothalamic Pathway"; Guo et al., 2017), corticothalamic feedback modulates not only MGBv but also A1 and TRN. Furthermore, corticothalamic feedback via TRN inhibiting MGBv seems to determine whether A1 neurons respond to weak tones or not (Ibrahim et al., 2021). Future investigations will need to explore the joint effects of the combined thalamocortical-corticothalamic-corticoreticular-intracortical loop. In addition, forward and feedback effects must be studied in the context of natural stimuli, such as communication sounds, and with the consideration of specific task goals and motivations.

Temporal Representation and Precision

MGBv neurons are highly sensitive to temporally fluctuating sound, such as amplitude modulated tones or noise. The majority of these neurons show strongly synchronized firing patterns to temporal modulations of 20–40 Hz, but phase-locking is typically

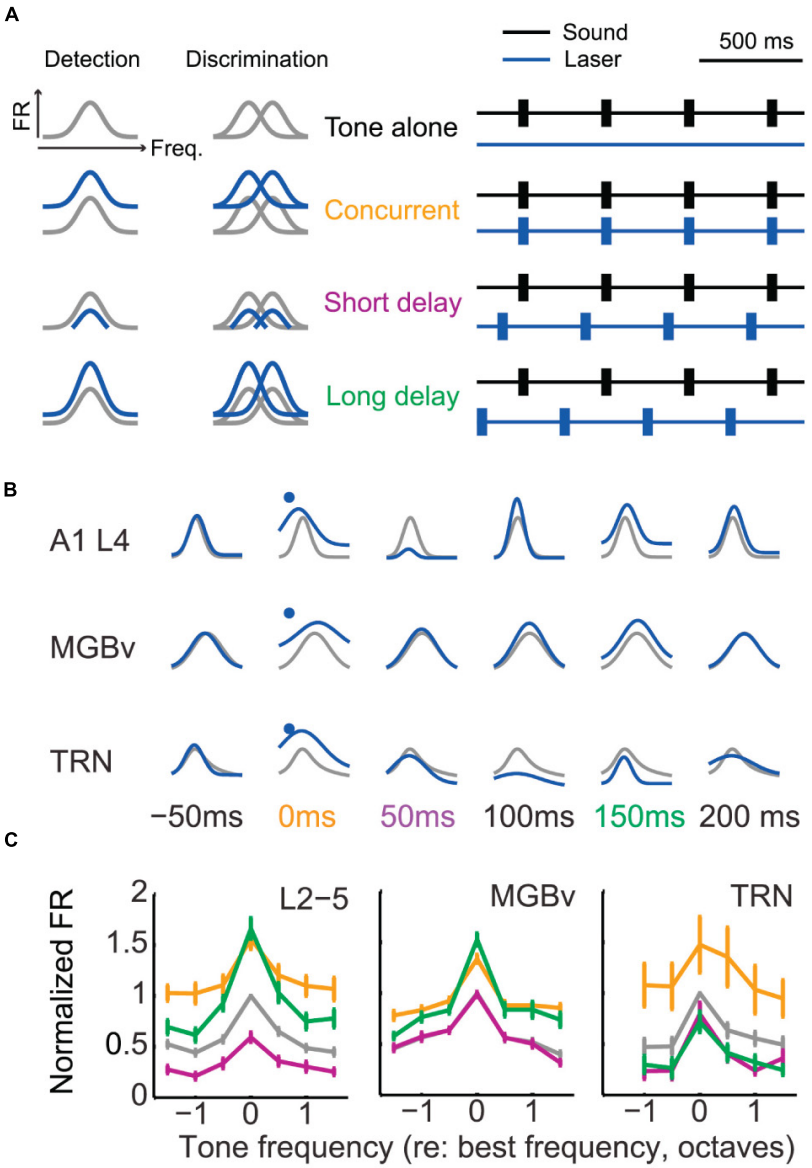

FIGURE 4 | Corticothalamic gain control on frequency tuning curves of A1, MGBv, and TRN neurons. **(A)** Schematic of corticothalamic modulation on the turning curves of A1 cortical neurons and optogenetic laser stimulation paradigm. Lemniscal layer VI corticothalamic neurons were activated by expressing ChR2 in A1 bilaterally in Ntsr1-Cre mice and using pulses of blue laser light. Depending on the duration of the sound stimulation delay following laser activation of corticothalamic neurons in layer VI, the tuning curves were predicted to be modulated distinctively and to define tone detection and discrimination behaviors. The animals were trained to detect or discriminate sounds using an avoidance task. **(B)** Representative modulation effect on the tuning curves of A1, MGBv, and TRN neurons (gray, tone-alone; blue, tone-and-laser). Sound-evoked responses enhanced in A1 and MGBv neurons for the concurrent stimulation of tones and laser (orange). For the tone presentation with a short or long delay following the corticothalamic stimulation (purple or green), corticothalamic modulation effects differed in A1, MGBv, and TRN neurons, showing enhancement for one station while showing suppression for the others. **(C)** Tone-evoked firing rates (mean ± SEM) were normalized and compared to the values evoked for the BF of the tone-alone condition. When tone presentation and corticothalamic stimulation were concurrent, all A1, MGBv, and TRN neurons increased firing rates (paired *t* test, $p < 0.05$). When tones were presented with a short delay following laser stimulation, the firing rates decreased in A1 and TRN neurons ($p < 0.05$) but no change was found in MGBv neurons ($p > 0.05$). For the long delayed condition, A1 and MGBv neurons increased firing rates ($p < 0.05$), whereas TRN neurons reduced it ($p < 0.05$) (Adapted and modified with permission from Guo et al., 2017; Figures 4B, 5D,E).

limited to modulation frequencies below ∼60 to 200 Hz (guinea pigs: Creutzfeldt et al., 1980, cats: Rouiller et al., 1981; Miller et al., 2002, non-human primates: Preuss and Müller-Preuss, 1990; Bartlett and Wang, 2007, 2011). Those values are two to four times higher for MGBv neurons than for A1 neurons (Miller et al., 2002; Bartlett and Wang, 2007) reflecting the progressively reduced ability of envelope synchronization along the ascending pathway (Joris et al., 2004). While ∼40% of MGBv neurons with BFs mostly >1.5 kHz show exclusively a rate code, another ∼40% of MGBv neurons show both synchronized responses

and increased firing in a single neuron dependent on repetition rates, and the proportion was five times larger than that for A1, suggesting that MGBv is a transition stage for this computation (Bartlett and Wang, 2007; but see Yin et al., 2011). For some thalamocortical transmissions, the temporal code is transformed to a rate code due to the synaptic interactions, in which excitatory and inhibitory inputs both inherit a temporal code from MGBv but the spiking response loses synchronization if they are in-phase (Bartlett and Wang, 2007; Gao and Wehr, 2015). Furthermore, higher auditory cortical fields encode temporal fluctuations predominantly in firing rate, with the exception of very low temporal modulation frequencies (Scott et al., 2011; Hullett et al., 2016; Johnson et al., 2020).

MGBv neurons respond strongly to species-specific and other animal vocalizations, and the response preferences are consistent with the values estimated by conventional artificial stimuli synthesized to represent acoustic properties contained in vocalizations (Creutzfeldt et al., 1980; Symmes et al., 1980; Philibert et al., 2005; Suta et al., 2007). Phase-locking to the F0 of harmonic vocalization is observed when neurons have the capacity of phase locking to pure tones in the range of the F0 (Wallace et al., 2007). The temporal spike pattern of MGBv neurons more closely matches the spectrogram of vocalizations and shows higher decoding performance than A1 neurons (Huetz et al., 2009; Souffi et al., 2020), suggesting that some information of the vocalization content is still present in a temporal code at the thalamic level.

Thalamocortical neurons have been shown to switch between two distinct firing modes: "burst" and "tonic" (Steriade et al., 1993; Sherman, 2001; Llinás and Steriade, 2006). When the principal cells are depolarized by incoming inputs and switch into "tonic mode," voltage-gated T-type Ca^{2+} channels become inactivated. In this mode, the firing patterns show a linear relationship to the input strength and spikes occur with high temporal precision (Mease et al., 2014; Hasse and Briggs, 2017); therefore, the "tonic mode" is more suitable for sound discrimination sound. After a period of depolarization, cells switch into "burst mode" by re-activation of inward Ca^{2+} current. In this mode, cells are hyperpolarized and prone to produce spikes with lowered threshold and less temporal precision. Thus, the input-output relationship is highly non-linear (McCormick and Feeser, 1990; Zhan et al., 1999) and is thought to be more helpful for sound detection (Hu et al., 1994; Bartlett and Smith, 1999; but see Massaux et al., 2004).

Stimulation of corticothalamic neurons in layer VI induces depolarization in MGBv neurons shortly after hyperpolarization (Bartlett and Smith, 2002; Yu et al., 2004), indicating that corticothalamic feedback can induce MGBv neurons to act in the "tonic mode" as has been seen in other sensory systems (Mease et al., 2014; but see Denman and Contreras, 2015). Furthermore, corticothalamic activation reduces adaptation to rapid repetitive stimulation as excitatory and inhibitory synaptic depression differs between the first-order visual/somatosensory thalamus and TRN (Mease et al., 2014; Crandall et al., 2015). In the auditory thalamus, corticothalamic activity seems to reduce adaptation to less-salient modulated noise since inactivation of corticothalamic neurons blocked the reduction (Kommajosyula et al., 2021). Therefore, corticothalamic neurons may be helpful for maintaining precise temporal responses in sequential or rapidly fluctuating sounds, which is characteristic of music and speech, and especially in less-salient sounds.

The assumption that corticothalamic neurons modulate temporal precision in MGBv is mainly based on intracellular recording of responses to pure tones or broadband noise. The likelihood of tonic and burst modes in the thalamus is affected by the different brain states of waking, sleep, attentiveness, and anesthesia (Weyand et al., 2001; Massaux et al., 2004; Gent et al., 2018). In the auditory thalamocortical system, burst mode is suppressed when spectrotemporally modulated broadband noise is presented compared to spontaneous or tone-driven activity (Miller and Schreiner, 2000). Further investigations are needed that use stimulus with more naturalistic modulations and behaving animals in order to dissect how the corticothalamic feedback affects the perception of naturally modulated sound, such as speech and music.

Overall, corticothalamic neurons in layer VI have potentially three major physiological functions in their effects on MGBv neurons: (i) refining the receptive field structure, (ii) modulating response gain, and (iii) controlling temporal precision, by regulating the balance of monosynaptic excitation and disynaptic inhibition. These three aspects are not operating independently, and the causes and effects of these interactions remain to be explored in more detail. This is relevant in the context of the next question, namely how corticothalamic modulations shape sound analysis and auditory perception.

EFFECTS OF CORTICOTHALAMIC FEEDBACK ON SOUND PERCEPTION

Frequency Analysis

Although some physiological functions of lemniscal auditory corticothalamic neurons have been gradually revealed in the past decades, it remains to be determined how the feedback affects hearing abilities in common, natural tasks. It has been technically challenging to selectively target corticothalamic neurons for recording and/or manipulations while, at the same time, measuring perceptual attributes in awake, behaving animals. Although a recent study using optogenetic phototagging in awake mice showed that layer VI corticothalamic neurons have narrower frequency tuning and higher selectivity of information flow compared to layer V corticofugal neurons (Williamson and Polley, 2019), the studies for corticothalamic modulations discussed above were largely based on recordings under anesthesia. It has been postulated that corticothalamic feedback is, in particular, required for more complex sound processing in behaving animals since ablation of auditory cortex revealed performance deficits in discrimination of frequency modulated tones but not for simple frequency tones (Ohl et al., 1999; Ono et al., 2006). However, none of these studies could dissect the separate roles of cortico-cortical vs. thalamo-cortical vs. cortico-thalamic contributions. More recently, layer specific electrical microstimulation showed that modulation of signal detection and cortical frequency

processing do appear to involve recurrent cortico–thalamo–cortical interactions (Happel et al., 2014; Saldeitis et al., 2021). As mentioned above (see section "Gain Changes"), optogenetic activation of corticothalamic neurons in layer VI modulated sound detection and frequency discrimination abilities of mice (Guo et al., 2017). It demonstrated that the physiological effects of layer VI corticothalamic feedback, such as refining tuning curves and controlling gain of MGBv and A1 neurons, can contribute to modulate frequency perception. Collectively, the present understanding is that layer VI corticothalamic neurons that project to MGBv neurons can affect spectral perception and are likely critical for perceptually demanding situations.

Harmonic Structure Analysis

Mistuning Detection

We hypothesized that corticothalamic feedback from layer VI is particularly important for complex sound processing and focused our study on harmonic structure. Harmonicity perception is potentially regulated by all the three physiological effects of layer VI corticothalamic feedback proposed above. Harmonicity is a strong grouping cue in speech/music perception and scene analysis. Sounds with a harmonic structure are typically perceived as one single entity associated with a specific pitch (Roberts and Bregman, 1991; Roberts and Bailey, 1993), despite containing many different frequency components. When background noise has harmonic structure, detection of foreground sound is improved (Deroche and Culling, 2011; Steinmetzger and Rosen, 2015; Guest and Oxenham, 2019). In order to explore roles of harmonicity, inharmonic sound stimuli have been generated in several different ways by perturbing a regular frequency interval of harmonic structure. It can be achieved by simply shifting all the harmonics to lower or higher frequencies to the same degree (i.e., the frequency interval no longer matches to F0 but regular), stretching out frequency intervals of harmonic components (i.e., each interval differs in the harmonic series), or randomly shifting each harmonic in a small degree. Since the former two cases preserve some levels of spectral regularity, this moderately contributes to fused perception (Roberts and Brunstrom, 1998, 2001). Using harmonic and randomly inharmonic synthetic vowels, harmonicity is shown to improve the segregation of concurrent vowels (Culling and Darwin, 1994; de Cheveigné, 1995; de Cheveigné et al., 1997). When inharmonicity is artificially introduced by jittering each harmonic of speech, the accurate segregation of speech from competing speech or speech-like noise is impaired (Popham et al., 2018).

One broadly studied inharmonic tone paradigm is a mistuned complex tone, which comprises a harmonic shifted to lower or higher frequency in an otherwise harmonic complex tone (i.e., "mistuning") (**Figure 5**). Specifically, the shifted component can be heard as standing out as a separate tone for low frequency harmonics or produce a sensation of roughness in the sound quality for high frequency harmonics (Moore et al., 1985, 1986; Hartmann et al., 1990). Perception of the shifted component as a separate tone is dependent on the frequencies of shifted components and degrades for higher F0s (Hartmann et al., 1990; Gockel and Carlyon, 2018). Based on the cochlear filtering model, a harmonic is assumed resolved when it falls in a single filter bank, while it is considered unresolved when several harmonics excite the same filter (Houtsma and Smurzynski, 1990; Shackleton and Carlyon, 1994). Thus, mistuning may be detected as a deviated spectral component in a spectral template of resolved harmonics ("harmonic template") that is expected to be a series of multiple integers of the F0 of a harmonic sound ("spectral cue") (Goldstein, 1973; Terhardt, 1974; Lin and Hartmann, 1998). Alternatively, the sensation of roughness or "beating," which is produced by an interaction of adjacent frequency components within the same cochlea filter, has been thought to help detect disruption of harmonicity (Assmann and Summerfield, 1994; Culling and Darwin, 1994; but see de Cheveigné, 1999). Detecting an inharmonic sound may also be assisted by the temporal excitation patterns synchronizing to the envelope fluctuations ("temporal cue") (Licklider, 1951; Meddis and Hewitt, 1991a,b). Both, spectral and temporal cues can be used for detecting a change of harmonicity.

In terms of harmonicity perception by animals, they have been found to be sensitive to harmonic structures (Kalluri et al., 2008) and can perform behavioral tasks for F0 judgment (Tomlinson and Schwarz, 1988; Walker et al., 2009; Osmanski et al., 2013) and mistuning detection (Lohr and Dooling, 1998; Klinge and Klump, 2009; Homma et al., 2016). Smaller animals are more likely to rely on temporal cues due to their generally broader cochlea filters (Sumner et al., 2018; Walker et al., 2019).

Neural Responses to Harmonic Complex Sounds

Neural responses can encode both spectral and temporal cues to harmonic and inharmonic complex sound (for review; Micheyl and Oxenham, 2010). At the auditory nerve, neurons respond to individual frequency components and can follow the F0 of complex tones with phase-locked responses up to ∼1 kHz (Horst et al., 1986; Cariani and Delgutte, 1996; Sinex et al., 2003; Cedolin and Delgutte, 2005). At higher stations of the central auditory system, temporal cues are degraded, and sound information is integrated. For harmonic sounds, neurons in IC and A1 increase firing rate when resolved harmonics are close to the neurons' best frequencies and both stations show some phase-locking to the envelope periodicity (F0) for unresolved harmonics (Schwarz and Tomlinson, 1990; Steinschneider et al., 1998; Fishman et al., 2013; Su and Delgutte, 2019, 2020). The phase-locking limits to the F0, however, decrease tenfold from IC to A1. For inharmonic sounds, the neurons show phase-locking to the fine structure of the envelope as well as the periodicity of interactions between mistuned and neighboring harmonics ("beating") (Sinex et al., 2002, 2005; Fishman and Steinschneider, 2010; Homma et al., 2017). While synchronized responses can be observed regardless of the distance between a mistuned harmonic and a frequency that MGBv or IC neurons are tuned to, changes of temporal patterns are weaker for A1 neurons when a mistuned harmonic is far away from a tuned frequency. In addition, the changes of firing rates occur to inharmonic sounds. The neurons increase their firing rates compared to the responses to the harmonic sound that has the same spectral components except the mistuned harmonic. Although a proportion of the neurons shows opposite decreasing trend, it results in enhanced

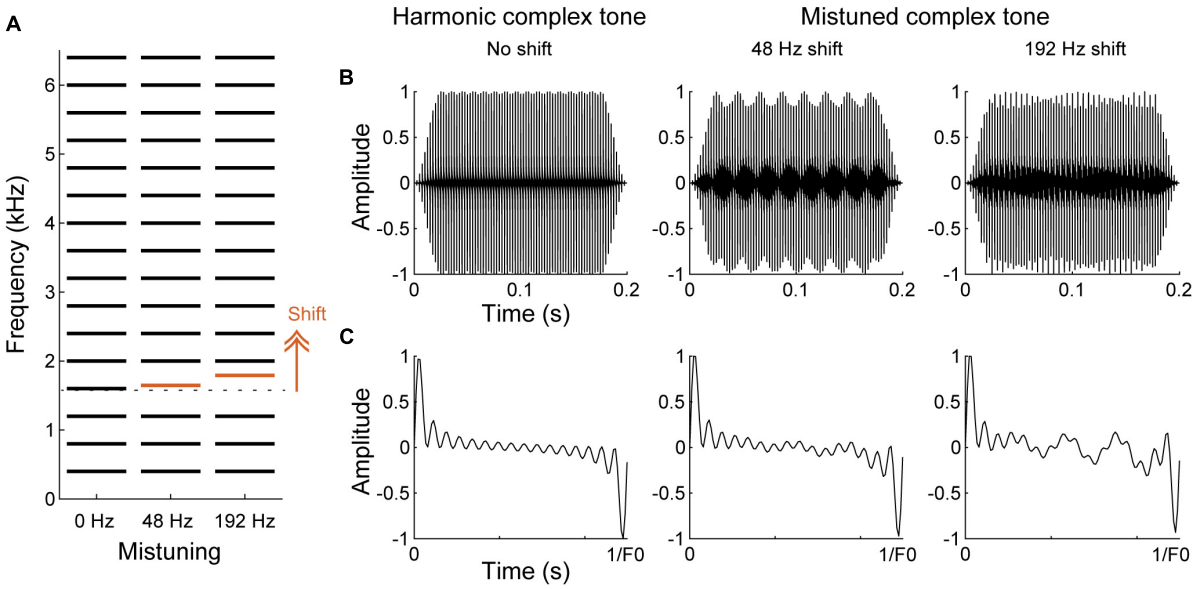

FIGURE 5 | Spectral and temporal cues in harmonic and mistuned complex tones. **(A)** Schematic of spectral structures of harmonic and mistuned complex tones comprised a F0 of 400 Hz with 16 harmonics. 4th harmonic (red) is shifted to higher frequencies for the mistuned complex tones by 48 or 192 Hz. **(B,C)** The waveforms of the complex tones shown in A for the overall tone duration **(B)** and for one cycle of the period of the harmonic complex tone **(C)**. Due to the mistuned harmonic, additional temporal fluctuations emerge for the mistuned complex tones.

responses for IC, MGBv and A1 neurons on average. Similar to temporal patterns, firing rates to mistuning are specifically enhanced in A1 when a mistuned harmonic is closer to a neuron's tuned-frequency, and the frequency specific changes of temporal patterns and firing rates in A1 are thought to correlate with "standing-out" perception of a mistuned harmonic in humans (Fishman and Steinschneider, 2010). It is unknown whether thalamocortical and/or corticothalamic projections contribute to form the specificity to the mistuned harmonic frequency in A1.

The corticothalamic modulation likely enhances spectral analysis of harmonicity in MGBv by sharpening the spectral tuning, i.e., increasing the sensitivity to small frequency shifts of mistuning. Furthermore, the corticothalamic modulation may assist detection of precise temporal excitation patterns of mistuning by switching to a "tonic mode" and improving temporal representation. Moreover, enhanced encoding of harmonic components and periodicities may refine the analysis of harmonic complex tones via spectral and temporal cues, respectively. Consequently, corticothalamic feedback may improve pitch discrimination.

In human and non-human primate auditory cortex, a population of neurons has been found to be specialized for harmonicity processing. Those neurons are excited by the F0 of a harmonic complex sound even when the actual F0 is omitted ("missing fundamental") or by periodic broadband noise stimulus that evoke pitch sensation, and were identified at the low frequency border of A1 and neighboring core region as well as at the adjacent belt regions (human: Patterson et al., 2002; Penagos et al., 2004; non-human primate: Bendor and Wang, 2005). For humans, non-primary auditory cortex is particularly critical for detecting pitch saliency and changes in pitch (Patterson et al., 2002; Penagos et al., 2004). Neurons responding to a subset of harmonics in harmonic complex tones ("harmonic template"), independent from their responses to pure tones, are scattered throughout the core regions of non-human primate (Feng and Wang, 2017). No equivalent type of response has been unequivocally identified at subcortical stations. It remains to be explored how the F0 representation emerges in the auditory system, whether the transformation is achieved by subcortical or cortical processing, and whether corticothalamic feedback is required.

Selective Elimination of Corticothalamic Neurons

We have demonstrated that selective elimination of layer VI corticothalamic neurons using chromophore targeted laser photolysis impairs the ability of ferrets to detect mistuned complex tones (Homma et al., 2017). In the study, ferrets were trained in a go/no-go task to detect an inharmonic tone, which comprises the mistuned 4th harmonic in an otherwise harmonic complex tone of 16 harmonics with a F0 of 400 Hz (**Figures 5**, **6A**). Then, fluorescent microbeads conjugated with a light-sensitive chromophore were injected bilaterally in MGBv, and >6 weeks later apoptosis was induced in the retrogradely labeled corticothalamic neurons in layer VI by focusing a infrared laser beam on A1 at the depth of the targeted layer (**Figure 6B**). About 60% of corticothalamic neurons were selectively eliminated. Mistuning sensitivity was measured behaviorally before and after the elimination of corticothalamic neurons, and the psychometric curve was constructed as a function of degree of mistuning (**Figure 6C**). Shifts of the 4th harmonic to a higher frequency ranged from 0 to 192 Hz. After the elimination of layer VI corticothalamic

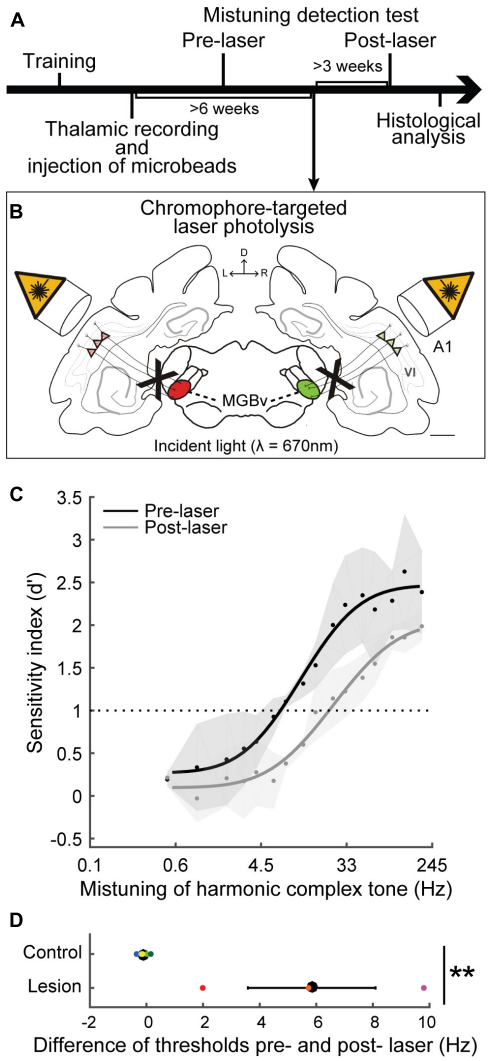

FIGURE 6 | Selective elimination of corticothalamic neurons in layer VI impaired mistuning detection performance. **(A)** Ferrets were trained for a mistuning detection task using go/no-go behavior paradigm. The fluorescent microbeads conjugated with a light-sensitive chromophore were injected into bilateral MGBv (left, red; right, green) guided by thalamic recordings. **(B)** Apoptosis was induced for the retrogradely labeled corticothalamic neurons in layer VI by infrared laser light on A1. D, dorsal; L, left; R, right; VI, cortical layer 6. Scale bar = 1 mm. **(C)** The psychometric curves were plotted as a function of degree of mistuning (on a log scale). X-axis indicates less mistuning at the left and larger mistuning toward the right. Mistuning detection performance was impaired, shifting the curve toward larger mistuning after selective elimination of corticothalamic neurons (gray) compared to the baseline (black). **(D)** The difference of threshold before and after laser illumination was larger for the animals that received corticothalamic elimination compared to the control animals (two-tailed unpaired *t* test, $^{**}p < 0.01$), supporting impaired behavior by a loss of A1-MGBv corticothalamic feedback. Individual animals are represented by colored dots (Adapted and modified from Homma et al., 2017; Figures 3D,F, 5B,C).

neurons, the psychometric functions were displaced to larger degrees of mistuning with reduced sensitivity (*d'*), indicating a deficit in discrimination ability. In addition, the threshold of detecting mistuning increased for the animals that received the corticothalamic lesion (**Figure 6D**). Although the lack of corticothalamic feedback could have reduced excitability of MGBv and decreased overall hearing sensitivity, there was no difference between lesion and control animals in the baseline performance of detecting an inharmonic tone with maximum degree of mistuning reinforced by a level difference of reference and target tones. These suggest A1-MGBv corticothalamic neurons are essential for successfully processing at least one important contributor to auditory scene analysis, namely for determining the harmonic structure of complex sounds. It remains to be elucidated how exactly the physiological changes of MGBv neurons by layer VI corticothalamic feedback contribute to mistuning detection related the temporal and spectral cues in mistuned complex tones. While reshaping of the receptive field structures is expected to improve spectral analysis of resolved harmonic components in MGBv, enhanced temporal precision in MGBv is predicted to refine temporal representation of periodicities in inharmonic complex tones. In addition, the modulation of gain titrates excitability of MGBv, which may control the focus to mistuned harmonic. All are plausible to increase the acuity of mistuning perception.

We postulate that corticothalamic feedback benefits other aspects of auditory scene analysis too. For example, enhanced signal detection by corticothalamic feedback seems to be robust for less salient inputs (Happel et al., 2014; Kommajosyula et al., 2021). Corticothalamic feedback may also contribute to signal-in-noise processing by amplifying weak foreground sounds by controlling the gain of MGBv neurons. Noise invariance also emerges from MGBv to A1 (Las et al., 2005; Rabinowitz et al., 2012; Schneider and Woolley, 2013; Souffi et al., 2020) and may be controlled by corticothalamic feedback. Finally, although speculative, the feedback may help in detecting onset synchrony, discriminating consonance and timbre, or assessing reverberation effects, by enhancing spectral and temporal processing in the MGBv.

SPEECH AND MUSIC PROCESSING IN MGBv

In this section, we will discuss potential roles of lemniscal corticothalamic feedback in speech and music processing. The ability of modulating frequency and harmonicity perception suggests layer VI corticothalamic projections can be involved in regulating speech and music recognition. Although we will mainly focus on MGBv, we briefly summarize findings in the human auditory cortex for an overall view of the auditory forebrain system (for review; Zatorre and Schönwiesner, 2011; Leonard and Chang, 2014). Growing evidence indicates that the left hemisphere is specialized for speech processing while the right hemisphere is dedicated for music processing (e.g., Zatorre et al., 1994; Griffiths et al., 1999; Tervaniemi et al., 2000; Albouy et al., 2020). This asymmetry is supported by finer temporal representation for speech in the left and superior spectral representation for music in the right (Zatorre et al., 2002; Poeppel, 2003). Furthermore, cortical regions are hierarchically

organized and functionally segregated. Spectrotemporal and phonological analyses for speech takes place in the left superior temporal gyrus (STG) and superior temporal sulcus (STS), which fall in the traditional "Wernicke's area" (Scott, 2000; Davis and Johnsrude, 2003; Mesgarani et al., 2014). Human voice identities are encoded in the right STS (Belin et al., 2000; Schall et al., 2015). While low-level sound features are mainly represented in the core regions, high-order sound features in speech and music are encoded in non-primary regions (Norman-Haignere et al., 2015). Although frequency information is first processed in the lemniscal core regions with their tonotopic organizations, non-primary regions play key roles in pitch and melody perception (Patterson et al., 2002; Penagos et al., 2004; Albouy et al., 2020). Stronger and more selective responses to a single speaker in competing simultaneous speech are observed in non-primary areas than in A1 (Ding and Simon, 2012; Mesgarani and Chang, 2012; O'Sullivan et al., 2019). Thus, the auditory cortex, particularly high-order regions, are essential for phonological, semantic, and melodic processing of speech and music.

Speech Processing

The human auditory thalamus is involved in some aspects of language processing. People with developmental dyslexia have difficulty in reading and writing and often exhibit impaired auditory and visual timing processing in various cortical and subcortical regions (for review; Ozernov-Palchik and Gaab, 2016; Stein, 2019). Some of those deficits could be traced to changes in the MGB (Galaburda et al., 1994; Diaz et al., 2012). Dyslexics often experience hindrances in auditory signal processing and sensorimotor processing. Psychoacoustic testing and auditory evoked potential studies show reduced sensitivities to discriminating temporally/spectrally modulated sound, or syllables (Stein and McAnally, 1995; Kraus et al., 1996; Menell et al., 1999; Goswami et al., 2002), and neuroimaging studies indicate deficits in rhythmic perception and audio-motor integration for dyslexics, which is in line with different neural phase alignment and consistency in the delta band compared to the control group (Hämäläinen et al., 2012; Colling et al., 2017). Functional magnetic resonance imaging (fMRI) revealed decreased activity in the left MGB for dyslexics in a syllabic discrimination task but not in passive listening condition (Diaz et al., 2012), supporting the possibility that top-down modulation, potentially corticothalamic feedback, differs between dyslexics and controls. Morphological examination showed that cells are smaller in the left than in the right MGB for dyslexics while no asymmetry is observed for control subjects (Galaburda et al., 1994). Dyslexics can exhibit malformation of cortical structures, "microgyria." An animal model expressing microgyria also exhibited abnormal anatomical changes in the MGB as well as temporal processing deficits in behavioral tasks similar to human dyslexics (Fitch et al., 1994; Herman et al., 1997; Peiffer et al., 2002; Anderson and Linden, 2016). Irregular connections between MGBv and microgyri may perturb temporal processing in dyslexics. Although further investigations are required of the anatomical and functional changes in MGB that may contribute to phonological skills, the ability to process syllables or words with high temporal precision for spoken language in MGB appears to be closely linked to processing of written language.

Other evidence that the auditory thalamus is involved in speech processing arises from the modulation of thalamic activity by cognitive demands during speech-based tasks via top-down feedback (Alain et al., 2005; Christensen et al., 2008; von Kriegstein et al., 2008; Mihai et al., 2019). Although speech representation in the human auditory cortex has been extensively studied (for review; Zatorre and Schönwiesner, 2011; Leonard and Chang, 2014), investigations in the human thalamus with non-invasive methods have been challenging due to its relatively small volume and the deep anatomical position in the brain. Thus, the thalamic activities measured by positron emission tomography (PET) (Salvi et al., 2002) or fMRI (Alain et al., 2005; Tervaniemi et al., 2006; Christensen et al., 2008), often only reflect overall responses of the thalamic complex to speech signal. Although language processing is lateralized to the left hemisphere in the cortex (Zatorre et al., 2002), the presence of an equivalent thalamic lateralization is not fully established. Contrasting activations to a consonant-vowel-consonant-vowel pseudoword were observed for a change of duration in left thalamus and for a change of frequency in right thalamus (Tervaniemi et al., 2006). In addition, only the activity in right thalamus significantly differed between diotic and dichotic attentive listening conditions while both sides were activated for one to three syllable nouns compared to reversed speech (Christensen et al., 2008). Two other studies support left lateralization for sentence and vowel processing in thalamus (Salvi et al., 2002; Alain et al., 2005). The left thalamus was activated for trials with successful identifications of two vowels that were concurrently presented (Alain et al., 2005) further supporting that the thalamus is involved in F0 discrimination with top-down modulation. It is, thus, likely that some basic speech processing aspects are lateralized to the left thalamus.

Recent fMRI studies with finer spatial resolution successfully identified MGB and captured the tonotopic organization in its ventral division (Moerel et al., 2015; Mihai et al., 2019). The activated responses for discriminating speech signal was observed in both sides of MGB, however, activity correlated to the behavioral performance in a speech recognition task was only observed in the left MGB (von Kriegstein et al., 2008; Mihai et al., 2019). Mihai et al. (2019) assigned two different attentional tasks while listening to an identical set of sound stimuli. They asked participants to report a change in either the presented syllables or the speaker identity (**Figure 7A**). Changes between syllables were reported for the speech task while detecting change of F0s was used for the speaker task. Although there was no significant difference between speech and speaker task for the blood-oxygen-level-dependent (BOLD) responses in MGBv, the correlation between the evoked activity and the behavior performance of the speech task was only found in the left MGBv (**Figures 7B–D**). This suggests that top-down modulations enhance speech processing in the left MGBv. Then, introducing speech-shaped white noise as background, Mihai et al. (2021) showed that the enhanced top-down modulation

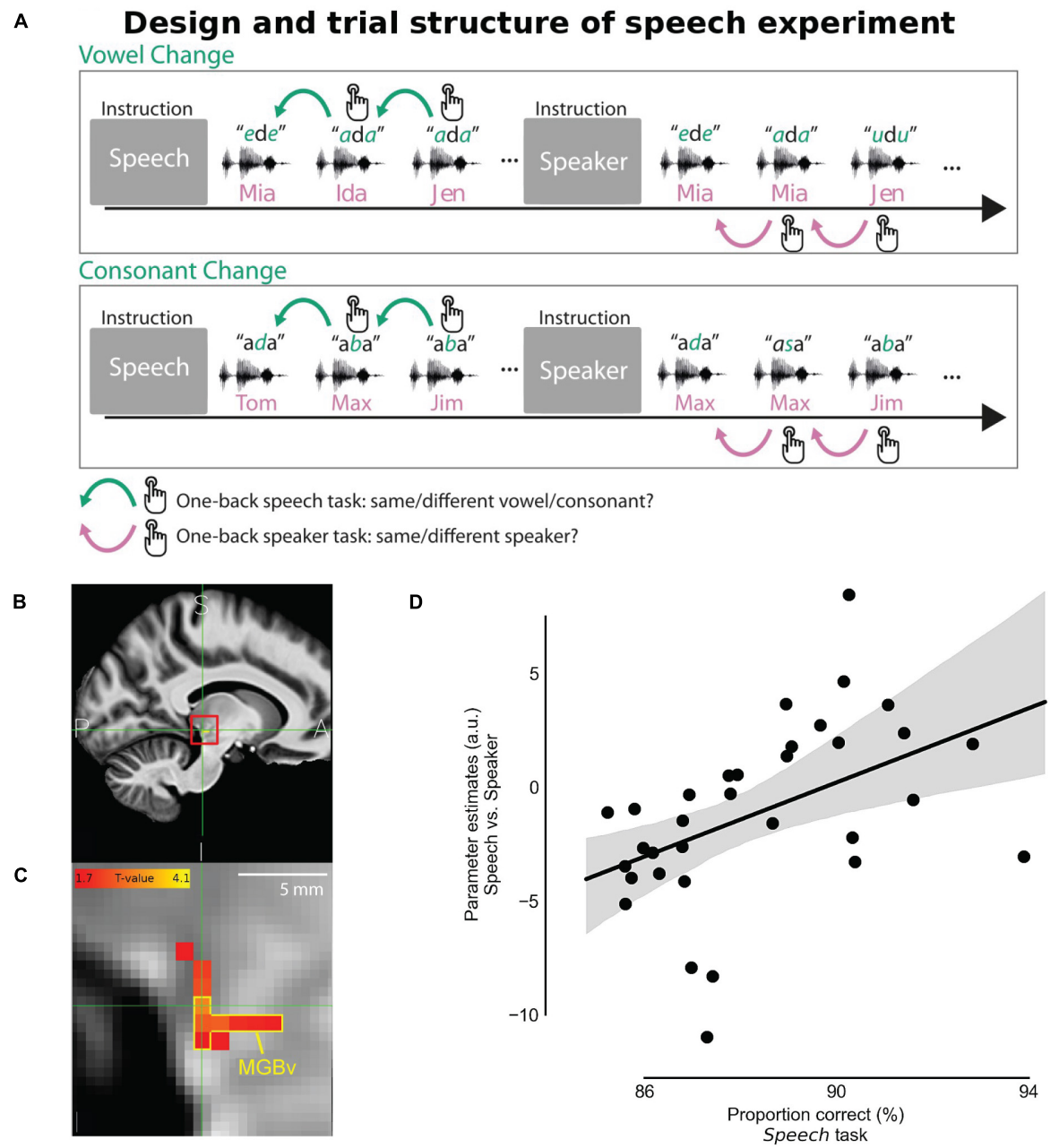

FIGURE 7 | Task-dependent modulation in MGBv for speech recognition. **(A)** Behavioral task design. In the speech task, participants pressed a button when a syllable changed in the sequence of vowel-consonant-vowel-syllables stimulus. In the speaker task, using exactly the same stimulus, they were instead asked to report when speaker identity changed in regardless of syllable changes. **(B)** The panel shows the averaged structural image by fMRI (sagittal section) across participants on the human brain atlas. The location of the left MGB was estimated and indicated with the red square. A, anterior; I, inferior; P, posterior; S, superior. **(C)** Zoomed view of the red square in the panel B, denoting MGBv by the yellow contour. The strength of the correlation between the speech vs. speaker task contrast and the averaged behavioral correct performance rate in the speech task was depicted with hot color coding. **(D)** Speech vs. Speaker tasks activation at the left MGBv coordinates correlated with the proportion of correct responses in the speech task. The better behavioral performance in the speech task corresponded to the larger difference of BOLD response between speech and speaker tasks in the left MGBv. Dots represent individual participants. The line shows the best fit with the gray area indicating 97% bootstrapped confidence interval (Adapted and modified with permission from Mihai et al., 2019; Figures 1C, 6, 7).

on speech recognition is strongly observed in left MGBv when listening condition is challenging.

Corticothalamic feedback potentially strengthens spectral and temporal processing in MGBv and modulates frequency and harmonicity perception. Thus, syllable discrimination may have relied on enhanced frequency tuning and more precise spike representation via the feedback modulation to characterize individual syllables. In particular, attention could

have modulated corticothalamic gain, affecting the excitability of MGBv neurons. Furthermore, although F0 discrimination did not show correlations with a lateralized change in activity in the study by Mihai et al. (2019), attentional modulation may be more relevant in harder discrimination tasks, such as segregation of competing simultaneous speech. We demonstrated that corticothalamic feedback was essential for processing spectral and/or temporal cues of the harmonic complex tones and detecting mistuning. A similar mechanism could assist speaker discrimination, mainly based on F0 discrimination using spectral structure and periodicity. F0 discrimination also helps to segregate foreground sounds from background "noise." Thus, top-down modulation via corticothalamic feedback may assist segregating simultaneously presented signals, which is an essential function of auditory scene analysis.

It would be interesting to examine what features of speech are extracted at the lemniscal thalamus and how the processing is modulated by attention or task demands. The temporal resolution of magnetoencephalography (MEG), electroencephalography (EEG), or electrocorticography (ECoG) (<10 msec) is finer than fMRI (<5 s), and recent studies showed that activity in deep subcortical structures can be detected by MEG (Müller et al., 2019; Pizzo et al., 2019). Technical advances in temporal and spatial resolutions are expected to dissect in greater detail human speech processing mechanisms along the auditory pathways from subcortical to cortical stations. In the human auditory cortex, attentional switching between low and high frequencies changes the activated locations of fMRI voxels corresponding to the attended frequencies in the primary auditory areas (Da Costa et al., 2013). In animal studies using extracellular recordings, task engagement increases or decreases responses to the behavioral target sound and reshapes the receptive field structures in A1 neurons compared to passive listening and dependent on task difficulty (Fritz et al., 2003; Atiani et al., 2009; Lee and Middlebrooks, 2011; Niwa et al., 2012; Schwartz and David, 2018). Attentive modulation was, however, larger in the belt/parabelt areas than in the core areas (Atiani et al., 2014; Niwa et al., 2015; Elgueda et al., 2019). Thus, corticothalamic feedback may be a gate of top-down modulations for complex cortical processing, supporting precise acoustic representations in A1 via the aforementioned physiological functions. Future studies are necessary to elucidate how corticothalamic projections contribute to modulate representations of sound signals in A1 and higher cortical fields as well as in MGBv.

Music Processing

Music processing studies in the human auditory system often involve a comparison between musicians and non-musicians since musical training is believed to induce plastic changes of structure and function in cortical and subcortical regions. Professional musicians tend to start receiving perceptual and motor training in their early childhood; therefore, structural and functional changes could reflect enhanced music processing. Plastic changes to musical training are indeed observed in the thalamus of musicians. Pianists have greater gray matter volume for the right thalamus (Vaquero et al., 2016), and drummers show increased oscillatory activities between the thalamus and premotor cortex/posterior parietal cortex (Krause et al., 2010). Furthermore, the thalamus is involved in musical imagery, i.e., an evoked sensation of music without external source. The right thalamus was activated for this music imagination phenomenon (Goycoolea et al., 2007) and melody recall (Zatorre et al., 1996). In addition, the ventral thalamus was activated during melody or sentence generation (Brown et al., 2006). Pleasant feelings associated with music listening have been shown to activate the thalamus, especially the mediodorsal thalamus, which regulates emotional processing (Blood and Zatorre, 2001; Klepzig et al., 2020). Contrasting to the left-hemispheric lateralization for speech processing, these studies support right-hemispheric lateralization for music processing.

It is, however, still not well investigated how MGBv processes musical signals and what are the main functional roles of corticothalamic feedback for music processing. For example, increased activation to urban noise, including music, were observed in the MGB of schizophrenic patients, supporting a role for the thalamus in sensory gating (Tregellas et al., 2009). Subcortical auditory structures, MGB and IC, with a strong corticofugal input, showed greater synchronization and responses to pieces of music compared to a scrambled version of the music or ripple noise, suggesting top-down modulation specific to music perception as opposed to basic sound perception (Abrams et al., 2013; Jiang et al., 2013). The thalamus was more highly activated to changes of musical chords or timbre (different musical instruments), when they deviated from an expected musical flow (Koelsch et al., 2002), again pointing a role of expectation or surprise in guiding corticothalamic feedback.

Musical chords of Western music consist of multiple harmonic complex tones (>2); in other words, combinations of pitches/harmonics. If two tones are separated by 1 octave, the frequency ratio is 2:1 ("unison," e.g., A, 220 Hz and 440 Hz), preserving harmonicity and resulting in sounding pleasant ("consonant"). A combination of two tones with the ratio of 3:2 ("perfect 5th," e.g., C and G) maintains the regularity of spectral components; therefore, it is consonant and usually evokes positive valence emotions. The ratio of 6:5 ("minor third," e.g., A and C) has imperfect consonance and association to sadness (a typical difference between major and minor cords). At the other extreme, if the frequency ratio of two tones is 16:15 ("minor second," e.g., C and C#), it sounds unpleasant ("dissonant") and can evoke all variety of negative valence emotions. Musicians showed refined representation of musical chords in the auditory brainstem response compared to non-musicians, suggesting that top-down modulation by corticofugal projections could optimize subcortical activities to efficiently process music (Lee et al., 2009). The sensation of roughness or beating has been thought to contribute to dissonance perception, and the difference of consonance and dissonance is reflected in the phase-locking in A1 to the frequency interactions of spectral components (Plomp and Levelt, 1965; Fishman et al., 2001). It seems, however, spectral regularity, i.e., harmonicity or periodicity, plays a key role while the judgment of pleasantness is also dependent on Western musical experience and cultural environment (Tramo et al., 2001; Bidelman and Krishnan, 2009; McDermott et al., 2010, 2016;

Bowling et al., 2018). A recent study suggests that harmonicity octave judgment of melodic lines, a sequence of tones, shares a similar mechanism to fused perception for octave harmonic structures (Demany et al., 2021). Given its apparent importance for perception of harmonic structure, corticothalamic feedback may help to discriminate musical chords and perceive musical melodies. Moreover, it may potentially help in the perception of rhythmic activity and its coordination between cortex and thalamus (Lee, 2013; Musacchia et al., 2014).

High-resolution human imaging and recording techniques are expected to shed light on the role of thalamocortical activity for a variety of music-based sound aspects in the near future.

Finally, speech and music processing are closely related and show overlaps of their functions. For example, musical training can improve language processing for children, adults or patients with language disorders (for review; Kraus and Slater, 2015; Coffey et al., 2017). Elementary school children (∼8 years old) who received musicianship classes, which included lessons of pitch and rhythm identification, and instrumental classes for 2 years showed better ability to correctly hear out speech from speech-shaped background noise than the controls that had only 1-year training (Slater et al., 2015). Musical training improves not only pitch encoding at subcortical and cortical levels for music and speech (Schön et al., 2004; Wong et al., 2007) but also speech-in-noise performance with enhanced temporal and spectral representation of speech (Parbery-Clark et al., 2011; Kraus et al., 2014; Swaminathan et al., 2015; Zendel et al., 2015). Since non-invasive techniques used in those studies did not explicitly identify which subcortical stations were involved, it remains to be examined what neural circuits are contributing for the improvement. As one of major corticofugal projections, layer VI A1-MGBv corticothalamic feedback could enhance spectral and temporal encoding of music and speech in the thalamus although other corticofugal connections from layer V to MGB and IC also may be critical for top-down modulations and plastic changes to musical training.

CLINICAL RELEVANCE

Along with musical training, corticothalamic feedback modulation may generally reinforce experience-dependent sound processing. Although speech processing ability degrades with aging, which is associated with impaired temporal coding and altered inhibitory signaling in the auditory system (Caspary et al., 2008; Gordon-Salant et al., 2011; Richardson et al., 2013; Presacco et al., 2016), temporal precision is actually improved in the MGB of aged-animals, which could indicate a compensation for the degraded hearing abilities via top-down modulation (Kommajosyula et al., 2019, 2021; Quraishe et al., 2020). Aged-musicians showed less degraded performance on the tasks that typically decline with aging (e.g., signal-in-noise, gap or mistuning detection) (Parbery-Clark et al., 2009; Zendel and Alain, 2012), which may reflect compensation via enhanced corticothalamic feedback. Thus, understanding corticofugal modulation may guide rehabilitation and training schemes for hearing impaired patients and therefore have clinical relevance.

CONCLUDING REMARKS

It has been proposed that corticothalamic neurons in layer VI (i) refine the receptive field of MGBv neurons, (ii) control gain of sound information flow, and (iii) increase temporal precision, by regulating the balance of excitation and inhibition. Lemniscal corticothalamic feedback can modulate the perception of frequency and harmonic structures, which is a basis for complex sound processing utilizing spectral and temporal cues and assisting auditory scene analysis of segregating concurrent speech or extracting signal from background noise. Task-related modulation was observed in human MGBv for speech processing particularly in noisy listening conditions. Although music processing in MGBv largely remains to be explored, corticothalamic feedback is expected to improve pitch perception and support musical appreciation. The future investigations will need to examine what aspects of speech and music corticothalamic feedback can modulate, but also to build up our understanding of the corticothalamic circuits including corticoreticular and intracortical pathways. The cutting-edge techniques of dissecting neural microcircuits in behaving animals and neuroimaging with finer spectral temporal resolutions in humans are expected to advance it. Ultimately, the better understanding of descending modulation may help improve rehabilitation for hearing impaired patients and musical training.

AUTHOR CONTRIBUTIONS

NH and VB both conceptualized the review. NH wrote the initial draft and prepared the figures. VB supported and reviewed the manuscript. Both authors made a substantial, direct and intellectual contribution to the review, edited and revised the manuscript, and approved the submitted version for publication.

FUNDING

This work was supported by the Japan Society for the Promotion of Science (Overseas Research Fellowship) to NH, by the Wellcome Trust UK (WT108369/Z/2015/Z), and by RNID funding to VB (S52_Bajo).

ACKNOWLEDGMENTS

We thank Christoph E. Schreiner and Andrew J. King for helpful discussions and comments during the preparation of the manuscript, the University of Iowa Musical Instrument Samples (http://theremin.music.uiowa.edu/MIS.html) for freely available tone samples of musical instruments, and the two reviewers for their helpful comments.

REFERENCES

Abrams, D. A., Ryali, S., Chen, T., Chordia, P., Khouzam, A., Levitin, D. J., et al. (2013). Inter-subject synchronization of brain responses during natural music listening. *Eur. J. Neurosci.* 37, 1458–1469. doi: 10.1111/ejn.12173

Aitkin, L. M., and Webster, W. R. (1972). Medial geniculate body of the cat: organization and responses to tonal stimuli of neurons in ventral division. *J. Neurophysiol.* 35, 365–380. doi: 10.1152/jn.1972.35.3.365

Alain, C., Reinke, K., McDonald, K. L., Chau, W., Tam, F., Pacurar, A., et al. (2005). Left thalamo-cortical network implicated in successful speech separation and identification. *Neuroimage* 26, 592–599. doi: 10.1016/j.neuroimage.2005.02.006

Albouy, P., Benjamin, L., Morillon, B., and Zatorre, R. J. (2020). Distinct sensitivity to spectrotemporal modulation supports brain asymmetry for speech and melody. *Science* 367, 1043–1047. doi: 10.1126/science.aaz3468

Andersen, R. A., Knight, P. L., and Merzenich, M. M. (1980a). The thalamocortical and corticothalamic conections of AI, AII, and the anteriior auditory field (AFF) in the cat: evidence ofr two largely sergregarted systems of connections. *J. Comp. Neurol.* 194, 663–701. doi: 10.1002/cne.901940312

Andersen, R. A., Roth, G. L., Aitkin, L. M., and Merzenich, M. M. (1980b). The efferent projections of the central nucleus and the pericentral nucleus of the inferior collculus in the cat. *J. Comp. Neurol.* 194, 649–662. doi: 10.1002/cne.901940311

Anderson, L. A., and Linden, J. F. (2011). Physiological differences between histologically defined subdivisions in the mouse auditory thalamus. *Hear. Res.* 274, 48–60. doi: 10.1016/j.heares.2010.12.016

Anderson, L. A., and Linden, J. F. (2016). Mind the Gap: two dissociable mechanisms of temporal processing in the auditory system. *J. Neurosci.* 36, 1977–1995. doi: 10.1523/JNEUROSCI.1652-15.2016

Arcelli, P., Frassoni, C., Regondi, M., Biasi, S. D., and Spreafico, R. (1997). GABAergic neurons in mammalian thalamus: A marker of thalamic complexity? *Brain Res. Bull.* 42, 27–37. doi: 10.1016/S0361-9230(96)00107-4

Arehart, K. H., Souza, P. E., Muralimanohar, R. K., and Miller, C. W. (2011). Effects of age on concurrent vowel perception in acoustic and simulated electroacoustic hearing. *J. Speech Lang. Hear. Res.* 54, 190–210.

Assmann, P. F., and Summerfield, Q. (1994). The contribution of waveform interactions to the perception of concurrent vowels. *J. Acoust. Soc. Am.* 95, 471–484. doi: 10.1121/1.408342

Atiani, S., David, S. V., Elgueda, D., Locastro, M., Radtke-Schuller, S., Shamma, S. A., et al. (2014). Emergent selectivity for task-relevant stimuli in higher-order auditory cortex. *Neuron* 82, 486–499. doi: 10.1016/j.neuron.2014.02.029

Atiani, S., Elhilali, M., David, S. V., Fritz, J. B., and Shamma, S. A. (2009). Task difficulty and performance induce diverse adaptive patterns in gain and shape of primary auditory cortical receptive fields. *Neuron* 61, 467–480. doi: 10.1016/j.neuron.2008.12.027

Atilgan, H., Town, S. M., Wood, K. C., Jones, G. P., Maddox, R. K., Lee, A. K. C., et al. (2018). Integration of visual information in auditory cortex promotes auditory scene analysis through multisensory binding. *Neuron* 97, 640–655.e4. doi: 10.1016/j.neuron.2017.12.034

Bajo, V. M., Nodal, F. R., Moore, D. R., and King, A. J. (2010). The descending corticocollicular pathway mediates learning-induced auditory plasticity. *Nat. Neurosci.* 13, 253–260. doi: 10.1038/nn.2466

Bajo, V. M., Rouiller, E. M., Welker, E., Clarke, S., Villa, A. E., de Ribaupierre, Y., et al. (1995). Morphology and spatial distribution of corticothalamic terminals originating from the cat auditory cortex. *Hear. Res.* 83, 161–174. doi: 10.1016/0378-5955(94)00199-Z

Bartlett, E., and Smith, P. (2002). Effects of paired-pulse and repetitive stimulation on neurons in the rat medial geniculate body. *Neuroscience* 113, 957–974. doi: 10.1016/S0306-4522(02)00240-3

Bartlett, E., Stark, J., Guillery, R., and Smith, P. (2000). Comparison of the fine structure of cortical and collicular terminals in the rat medial geniculate body. *Neuroscience* 100, 811–828. doi: 10.1016/S0306-4522(00)00340-7

Bartlett, E. L. (2013). The organization and physiology of the auditory thalamus and its role in processing acoustic features important for speech perception. *Brain Lang.* 126, 29–48. doi: 10.1016/j.bandl.2013.03.003

Bartlett, E. L., Sadagopan, S., and Wang, X. (2011). Fine frequency tuning in monkey auditory cortex and thalamus. *J. Neurophysiol.* 106, 849–859. doi: 10.1152/jn.00559.2010

Bartlett, E. L., and Smith, P. H. (1999). Anatomic, intrinsic, and synaptic properties of dorsal and ventral division neurons in rat medial geniculate body. *J. Neurophysiol.* 81, 1999–2016. doi: 10.1152/jn.1999.81.5.1999

Bartlett, E. L., and Wang, X. (2007). Neural representations of temporally modulated signals in the auditory thalamus of awake primates. *J. Neurophysiol.* 97, 1005–1017. doi: 10.1152/jn.00593.2006

Bartlett, E. L., and Wang, X. (2011). Correlation of neural response properties with auditory thalamus subdivisions in the awake marmoset. *J. Neurophysiol.* 105, 2647–2667. doi: 10.1152/jn.00238.2010

Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., and Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature* 403, 309–312. doi: 10.1038/35002078

Bendor, D., and Wang, X. (2005). The neuronal representation of pitch in primate auditory cortex. *Nature* 436, 1161–1165. doi: 10.1038/nature03867

Bidelman, G. M., and Krishnan, A. (2009). Neural correlates of consonance, dissonance, and the hierarchy of musical pitch in the human brainstem. *J. Neurosci.* 29, 13165–13171. doi: 10.1523/JNEUROSCI.3900-09.2009

Blood, A. J., and Zatorre, R. J. (2001). Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion. *Proc. Natl. Acad. Sci. U.S.A.* 98, 11818–11823. doi: 10.1073/pnas.191355898

Bordi, F., and LeDoux, J. E. (1994). Response properties of single units in areas of rat auditory thalamus that project to the amygdala - II. Cells receiving convergent auditory and somatosensory inputs and cells antidromically activated by amygdala stimulation. *Exp. Brain Res.* 98, 275–286. doi: 10.1007/BF00228415

Bortone, D. S., Olsen, S. R., and Scanziani, M. (2014). Translaminar inhibitory cells recruited by layer 6 corticothalamic neurons suppress visual cortex. *Neuron* 82, 474–485. doi: 10.1016/j.neuron.2014.02.021

Bowling, D. L., Purves, D., and Gill, K. Z. (2018). Vocal similarity predicts the relative attraction of musical chords. *Proc. Natl. Acad. Sci. U.S.A.* 115, 216–221. doi: 10.1073/pnas.1713206115

Bregman, A. (1990). *Auditory Scene Analysis: The Perceptual Organization of Sound*. Cambridge, MA: The MIT Press.

Bregman, A. S., and Campbell, J. (1971). Primary auditory stream segregation and perception of order in rapid sequences of tones. *J. Exp. Psychol.* 89, 244–249. doi: 10.1037/h0031163

Brown, S., Martinez, M. J., and Parsons, L. M. (2006). Music and language side by side in the brain: a PET study of the generation of melodies and sentences. *Eur. J. Neurosci.* 23, 2791–2803. doi: 10.1111/j.1460-9568.2006.04785.x

Budinger, E., Brosch, M., Scheich, H., and Mylius, J. (2013). The subcortical auditory structures in the mongolian gerbil: II. Frequency-related topography of the connections with cortical field AI. *J. Comp. Neurol.* 521, 2772–2797. doi: 10.1002/cne.23314

Calford, M. (1983). The parcellation of the medial geniculate body of the cat defined by the auditory response properties of single units. *J. Neurosci.* 3, 2350–2364. doi: 10.1523/JNEUROSCI.03-11-02350.1983

Calford, M., and Aitkin, L. (1983). Ascending projections to the medial geniculate body of the cat: evidence for multiple, parallel auditory pathways through thalamus. *J. Neurosci.* 3, 2365–2380. doi: 10.1523/JNEUROSCI.03-11-02365.1983

Cariani, P. A., and Delgutte, B. (1996). Neural correlates of the pitch of complex tones. I. Pitch and pitch salience. *J. Neurophysiol.* 76, 1698–1716. doi: 10.1152/jn.1996.76.3.1698

Caspary, D. M., Ling, L., Turner, J. G., and Hughes, L. F. (2008). Inhibitory neurotransmission, plasticity and aging in the mammalian central auditory system. *J. Exp. Biol.* 211, 1781–1791. doi: 10.1242/jeb.013581

Cedolin, L., and Delgutte, B. (2005). Pitch of complex tones: rate-place and interspike interval representations in the auditory nerve. *J. Neurophysiol.* 94, 347–362. doi: 10.1152/jn.01114.2004

Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *J. Acoust. Soc. Am.* 25, 975–979. doi: 10.1121/1.1907229

Chi, T., Gao, Y., Guyton, M. C., Ru, P., and Shamma, S. (1999). Spectro-temporal modulation transfer functions and speech intelligibility. *J. Acoust. Soc. Am.* 106, 2719–2732. doi: 10.1121/1.428100

Christensen, T. A., Antonucci, S. M., Lockwood, J. L., Kittleson, M., and Plante, E. (2008). Cortical and subcortical contributions to the attentive processing of speech. *Neuroreport* 19, 1101–1105. doi: 10.1097/WNR.0b013e3283060a9d

Clarke, B. A., and Lee, C. C. (2018). Inhibitory projections in the mouse auditory tectothalamic system. *Brain Sci.* 8:103. doi: 10.3390/brainsci8060103

Clayton, K. K., Williamson, R. S., Hancock, K. E., Tasaka, G., Mizrahi, A., Hackett, T. A., et al. (2021). Auditory corticothalamic neurons are recruited by motor preparatory inputs. *Curr. Biol.* 31, 310–321.e5. doi: 10.1016/j.cub.2020.10.027

Coffey, E. B. J., Mogilever, N. B., and Zatorre, R. J. (2017). Speech-in-noise perception in musicians: a review. *Hear. Res.* 352, 49–69. doi: 10.1016/j.heares.2017.02.006

Colling, L. J., Noble, H. L., and Goswami, U. (2017). Neural entrainment and sensorimotor synchronization to the beat in children with developmental dyslexia: an EEG study. *Front. Neurosci.* 11:360. doi: 10.3389/fnins.2017.00360

Conley, M., Kupersmith, A. C., and Diamond, I. T. (1991). The organization of projections from subdivisions of the auditory cortex and thalamus to the auditory sector of the thalamic reticular nucleus in Galago. *Eur. J. Neurosci.* 3, 1089–1103. doi: 10.1111/j.1460-9568.1991.tb00044.x

Cotillon-Williams, N., Huetz, C., Hennevin, E., and Edeline, J.-M. (2008). Tonotopic control of auditory thalamus frequency tuning by reticular thalamic neurons. *J. Neurophysiol.* 99, 1137–1151. doi: 10.1152/jn.01159.2007

Crandall, S. R., Cruikshank, S. J., and Connors, B. W. (2015). A corticothalamic switch: controlling the thalamus with dynamic synapses. *Neuron* 86, 768–782. doi: 10.1016/j.neuron.2015.03.040

Creutzfeldt, O., Hellweg, F. C., and Schreiner, C. (1980). Thalamocortical transformation of responses to complex auditory stimuli. *Exp. Brain Res.* 39, 87–104. doi: 10.1007/BF00237072

Culling, J. F., and Darwin, C. J. (1993). Perceptual separation of simultaneous vowels: Within and across-formant grouping by F 0. *J. Acoust. Soc. Am.* 93, 3454–3467. doi: 10.1121/1.405675

Culling, J. F., and Darwin, C. J. (1994). Perceptual and computational separation of simultaneous vowels: cues arising from low-frequency beating. *J. Acoust. Soc. Am.* 95, 1559–1569. doi: 10.1121/1.408543

Da Costa, S., van der Zwaag, W., Miller, L. M., Clarke, S., and Saenz, M. (2013). Tuning in to sound: frequency-selective attentional filter in human primary auditory cortex. *J. Neurosci.* 33, 1858–1863. doi: 10.1523/JNEUROSCI.4405-12.2013

Darwin, C. J. (1984). Perceiving vowels in the presence of another sound: constraints on formant perception. *J. Acoust. Soc. Am.* 76, 1636–1647. doi: 10.1121/1.391610

Darwin, C. J., and Ciocca, V. (1992). Grouping in pitch perception: effects of onset asynchrony and ear of presentation of a mistuned component. *J. Acoust. Soc. Am.* 91, 3381–3390. doi: 10.1121/1.402828

Davis, M. H., and Johnsrude, I. S. (2003). Hierarchical processing in spoken language comprehension. *J. Neurosci.* 23, 3423–3431. doi: 10.1523/JNEUROSCI.23-08-03423.2003

de Cheveigné, A. (1999). Waveform interactions and the segregation of concurrent vowels. *J. Acoust. Soc. Am.* 106, 2959–2972. doi: 10.1121/1.428115

de Cheveigné, A. (1995). Identification of concurrent harmonic and inharmonic vowels: a test of the theory of harmonic cancellation and enhancement. *J. Acoust. Soc. Am.* 97, 3736–3748. doi: 10.1121/1.412389

de Cheveigné, A., McAdams, S., and Marin, C. M. H. (1997). Concurrent vowel identification. II. Effects of phase, harmonicity, and task. *J. Acoust. Soc. Am.* 101, 2848–2856. doi: 10.1121/1.419476

De La Mothe, L. A., Blumell, S., Kajikawa, Y., and Hackett, T. A. (2006). Thalamic connections of the auditory cortex in marmoset monkeys: core and medial belt regions. *J. Comp. Neurol.* 496, 72–96. doi: 10.1002/cne.20924

Demany, L., Monteiro, G., Semal, C., Shamma, S., and Carlyon, R. P. (2021). The perception of octave pitch affinity and harmonic fusion have a common origin. *Hear. Res.* 404:108213. doi: 10.1016/j.heares.2021.108213

Denman, D. J., and Contreras, D. (2015). Complex effects on *in vivo* visual responses by specific projections from mouse cortical layer 6 to dorsal lateral Geniculate Nucleus. *J. Neurosci.* 35, 9265–9280. doi: 10.1523/JNEUROSCI.0027-15.2015

Deroche, M. L. D., and Culling, J. F. (2011). Narrow noise band detection in a complex masker: masking level difference due to harmonicity. *Hear. Res.* 282, 225–235. doi: 10.1016/j.heares.2011.07.005

Diaz, B., Hintz, F., Kiebel, S. J., and von Kriegstein, K. (2012). Dysfunction of the auditory thalamus in developmental dyslexia. *Proc. Natl. Acad. Sci. U.S.A.* 109, 13841–13846. doi: 10.1073/pnas.1119828109

Ding, N., Patel, A. D., Chen, L., Butler, H., Luo, C., and Poeppel, D. (2017). Temporal modulations in speech and music. *Neurosci. Biobehav. Rev.* 81, 181–187. doi: 10.1016/j.neubiorev.2017.02.011

Ding, N., and Simon, J. Z. (2012). Emergence of neural encoding of auditory objects while listening to competing speakers. *Proc. Natl. Acad. Sci.U.S.A.* 109, 11854–11859. doi: 10.1073/pnas.1205381109

Drullman, R., Festen, J. M., and Plomp, R. (1994). Effect of temporal envelope smearing on speech reception. *J. Acoust. Soc. Am.* 95, 1053–1064. doi: 10.1121/1.408467

Edeline, J. M., Manunta, Y., Nodal, F. R., and Bajo, V. M. (1999). Do auditory responses recorded from awake animals reflect the anatomical parcellation of the auditory thalamus? *Hear. Res.* 131, 135–152. doi: 10.1016/s0378-5955(99)00026-x

Edeline, J.-M., and Weinberger, N. M. (1991). Thalamic short-term plasticity in the auditory system: associative retuning of receptive fields in the ventral medial geniculate body. *Behav. Neurosci.* 105, 618–639. doi: 10.1037/0735-7044.105.5.618

Elgueda, D., Duque, D., Radtke-Schuller, S., Yin, P., David, S. V., Shamma, S. A., et al. (2019). State-dependent encoding of sound and behavioral meaning in a tertiary region of the ferret auditory cortex. *Nat. Neurosci.* 22, 447–459. doi: 10.1038/s41593-018-0317-8

Elliott, T. M., and Theunissen, F. E. (2009). The modulation transfer function for speech intelligibility. *PLoS Comput. Biol.* 5:e1000302. doi: 10.1371/journal.pcbi.1000302

Eramudugolla, R., McAnally, K. I., Martin, R. L., Irvine, D. R. F., and Mattingley, J. B. (2008). The role of spatial location in auditory search. *Hear. Res.* 238, 139–146. doi: 10.1016/j.heares.2007.10.004

Feng, L., and Wang, X. (2017). Harmonic template neurons in primate auditory cortex underlying complex sound processing. *Proc. Natl. Acad. Sci. U.S.A.* 114, E840–E848. doi: 10.1073/pnas.1607519114

Fishman, Y. I., Micheyl, C., and Steinschneider, M. (2013). Neural representation of harmonic complex tones in primary auditory cortex of the awake monkey. *J. Neurosci.* 33, 10312–10323. doi: 10.1523/JNEUROSCI.0020-13.2013

Fishman, Y. I., and Steinschneider, M. (2010). Neural correlates of auditory scene analysis based on inharmonicity in monkey primary auditory cortex. *J. Neurosci.* 30, 12480–12494. doi: 10.1523/JNEUROSCI.1780-10.2010

Fishman, Y. I., Volkov, I. O., Noh, M. D., Garell, P. C., Bakken, H., Arezzo, J. C., et al. (2001). Consonance and dissonance of musical chords: neural correlates in auditory cortex of monkeys and humans. *J. Neurophysiol.* 86, 2761–2788. doi: 10.1152/jn.2001.86.6.2761

Fitch, R. H., Tallal, P., Brown, C. P., Galaburda, A. M., and Rosen, G. D. (1994). Induced microgyria and auditory temporal processing in rats: A model for language impairment? *Cereb. Cortex* 4, 260–270. doi: 10.1093/cercor/4.3.260

Fritz, J., Shamma, S., Elhilali, M., and Klein, D. (2003). Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex. *Nat. Neurosci.* 6, 1216–1223. doi: 10.1038/nn1141

Galaburda, A. M., Menard, M. T., and Rosen, G. D. (1994). Evidence for aberrant auditory anatomy in developmental dyslexia. *Proc. Natl. Acad. Sci. U.S.A.* 91, 8010–8013. doi: 10.1073/pnas.91.17.8010

Gao, X., and Wehr, M. (2015). A coding transformation for temporally structured sounds within auditory cortical neurons. *Neuron* 86, 292–303. doi: 10.1016/j.neuron.2015.03.004

Gent, T. C., Bandarabadi, M., Herrera, C. G., and Adamantidis, A. R. (2018). Thalamic dual control of sleep and wakefulness. *Nat. Neurosci.* 21, 974–984. doi: 10.1038/s41593-018-0164-7

Gockel, H. E., and Carlyon, R. P. (2018). Detection of mistuning in harmonic complex tones at high frequencies. *Acta Acust. United Acust.* 104, 766–769. doi: 10.3813/AAA.919219

Goldstein, J. L. (1973). An optimum processor theory for the central formation of the pitch of complex tones. *J. Acoust. Soc. Am.* 54, 1496–1516. doi: 10.1121/1.1978261

Gordon-Salant, S., Fitzgibbons, P. J., and Yeni-Komshian, G. H. (2011). Auditory temporal processing and aging: implications for speech understanding of older people. *Audiol. Res.* 1:e4. doi: 10.4081/audiores.2011.e4

Goswami, U., Thomson, J., Richardson, U., Stainthorp, R., Hughes, D., Rosen, S., et al. (2002). Amplitude envelope onsets and developmental dyslexia: a new hypothesis. *Proc. Natl. Acad. Sci. U.S.A.* 99, 10911–10916. doi: 10.1073/pnas.122368599

Goycoolea, M. V., Mena, I., Neubauer, S. G., Levy, R. G., Fernández Grez, M., and Berger, C. G. (2007). Musical brains: a study of spontaneous and evoked musical

sensations without external auditory stimuli. *Acta Otolaryngol.* 127, 711–721. doi: 10.1080/00016480601053057

Greenberg, S., Carvey, H., Hitchcock, L., and Chang, S. (2003). Temporal properties of spontaneous speech-a syllable-centric perspective. *J. Phon.* 31, 465–485. doi: 10.1016/j.wocn.2003.09.005

Griffiths, T. D., Johnsrude, I., Dean, J. L., and Green, G. G. R. (1999). A common neural substrate for the analysis of pitch and duration pattern in segmented sound? *Neuroreport* 10, 3825–3830. doi: 10.1097/00001756-199912160-00019

Guest, D. R., and Oxenham, A. J. (2019). The role of pitch and harmonic cancellation when listening to speech in harmonic background sounds. *J. Acoust. Soc. Am.* 145, 3011–3023. doi: 10.1121/1.5102169

Guo, W., Clause, A. R., Barth-Maron, A., and Polley, D. B. (2017). A corticothalamic circuit for dynamic switching between feature detection and discrimination. *Neuron* 95, 180–194.e5. doi: 10.1016/j.neuron.2017.05.019

Hackett, T. A., Barkat, T. R., O'Brien, B. M. J., Hensch, T. K., and Polley, D. B. (2011). Linking topography to tonotopy in the mouse auditory thalamocortical circuit. *J. Neurosci.* 31, 2983–2995. doi: 10.1523/JNEUROSCI.5333-10.2011

Hämäläinen, J. A., Rupp, A., Soltész, F., Szücs, D., and Goswami, U. (2012). Reduced phase locking to slow amplitude modulation in adults with dyslexia: an MEG study. *Neuroimage* 59, 2952–2961. doi: 10.1016/j.neuroimage.2011.09.075

Happel, M. F. K., Deliano, M., Handschuh, J., and Ohl, F. W. (2014). Dopamine-modulated recurrent corticoefferent feedback in primary sensory cortex promotes detection of behaviorally relevant stimuli. *J. Neurosci.* 34, 1234–1247. doi: 10.1523/JNEUROSCI.1990-13.2014

Hartmann, W. M., McAdams, S., and Smith, B. K. (1990). Hearing a mistuned harmonic in an otherwise periodic complex tone. *J. Acoust. Soc. Am.* 88, 1712–1724. doi: 10.1121/1.400246

Hasse, J. M., and Briggs, F. (2017). Corticogeniculate feedback sharpens the temporal precision and spatial resolution of visual signals in the ferret. *Proc. Natl. Acad. Sci. U.S.A.* 114, E6222–E6230. doi: 10.1073/pnas.1704524114

He, J. (1997). Modulatory effects of regional cortical activation on the onset responses of the cat medial geniculate neurons. *J. Neurophysiol.* 77, 896–908. doi: 10.1152/jn.1997.77.2.896

He, J. (2003). Corticofugal modulation on both ON and OFF responses in the nonlemniscal auditory thalamus of the guinea pig. *J. Neurophysiol.* 89, 367–381. doi: 10.1152/jn.00593.2002

He, J., Yu, Y.-Q., Xiong, Y., Hashikawa, T., and Chan, Y.-S. (2002). Modulatory effect of cortical activation on the lemniscal auditory thalamus of the Guinea pig. *J. Neurophysiol.* 88, 1040–1050. doi: 10.1152/jn.2002.88.2.1040

Herman, A. E., Galaburda, A. M., Fitch, R. H., Carter, A. R., and Rosen, G. D. (1997). Cerebral microgyria, thalamic cell size and auditory temporal processing in male and female rats. *Cereb. Cortex* 7, 453–464. doi: 10.1093/cercor/7.5.453

Holdgraf, C. R., de Heer, W., Pasley, B., Rieger, J., Crone, N., Lin, J. J., et al. (2016). Rapid tuning shifts in human auditory cortex enhance speech intelligibility. *Nat. Commun.* 7:13654. doi: 10.1038/ncomms13654

Holstege, G., and Collewijn, H. (1982). The efferent connections of the nucleus of the optic tract and the superior colliculus in the rabbit. *J. Comp. Neurol.* 209, 139–175. doi: 10.1002/cne.902090204

Homma, N. Y., Bajo, V. M., Happel, M. F. K., Nodal, F. R., and King, A. J. (2016). Mistuning detection performance of ferrets in a go/no-go task. *J. Acoust. Soc. Am.* 139, EL246–EL251. doi: 10.1121/1.4954378

Homma, N. Y., Happel, M. F. K., Nodal, F. R., Ohl, F. W., King, A. J., and Bajo, V. M. (2017). A role for auditory corticothalamic feedback in the perception of complex sounds. *J. Neurosci.* 37, 6149–6161. doi: 10.1523/JNEUROSCI.0397-17.2017

Homma, N. Y., Hullett, P. W., Atencio, C. A., and Schreiner, C. E. (2020). Auditory cortical plasticity dependent on environmental noise statistics. *Cell Rep.* 30, 4445–4458.e5. doi: 10.1016/j.celrep.2020.03.014

Horst, J. W., Javel, E., and Farley, G. R. (1986). Coding of spectral fine structure in the auditory nerve. I. Fourier analysis of period and interspike interval histograms. *J. Acoust. Soc. Am.* 79, 398–416. doi: 10.1121/1.393528

Houtsma, A. J. M., and Smurzynski, J. (1990). Pitch identification and discrimination for complex tones with many harmonics. *J. Acoust. Soc. Am.* 87, 304–310. doi: 10.1121/1.399297

Hu, B., Senatorov, V., and Mooney, D. (1994). Lemniscal and non-lemniscal synaptic transmission in rat auditory thalamus. *J. Physiol.* 479, 217–231. doi: 10.1113/jphysiol.1994.sp020290

Huang, C. L., and Winer, J. A. (2000). Auditory thalamocortical projections in the cat: laminar and areal patterns of input. *J. Comp. Neurol.* 427, 302–331.

Huetz, C., Philibert, B., and Edeline, J.-M. (2009). A spike-timing code for discriminating conspecific vocalizations in the thalamocortical system of anesthetized and awake guinea pigs. *J. Neurosci.* 29, 334–350. doi: 10.1523/JNEUROSCI.3269-08.2009

Hullett, P. W., Hamilton, L. S., Mesgarani, N., Schreiner, C. E., and Chang, E. F. (2016). Human superior temporal Gyrus Organization of spectrotemporal modulation tuning derived from speech stimuli. *J. Neurosci.* 36, 2014–2026. doi: 10.1523/JNEUROSCI.1779-15.2016

Ibrahim, B. A., Murphy, C. A., Yudintsev, G., Shinagawa, Y., Banks, M. I., and Llano, D. A. (2021). Corticothalamic gating of population auditory thalamocortical transmission in mouse. *eLife* 10:e56645. doi: 10.7554/eLife.56645

Jiang, F., Stecker, G. C., and Fine, I. (2013). Functional localization of the auditory thalamus in individual human subjects. *Neuroimage* 78, 295–304. doi: 10.1016/j.neuroimage.2013.04.035

Johnson, J. S., Niwa, M., O'Connor, K. N., and Sutter, M. L. (2020). Amplitude modulation encoding in the auditory cortex: comparisons between the primary and middle lateral belt regions. *J. Neurophysiol.* 124, 1706–1726. doi: 10.1152/jn.00171.2020

Jones, E. G. (2007). *The Thalamus*, 2nd Edn. Cambridge: Cambridge University Press.

Joris, P. X., Schreiner, C. E., and Rees, A. (2004). Neural processing of amplitude-modulated sounds. *Physiol. Rev.* 84, 541–577. doi: 10.1152/physrev.00029.2003

Kalluri, S., Depireux, D. A., and Shamma, S. A. (2008). Perception and cortical neural coding of harmonic fusion in ferrets. *J. Acoust. Soc. Am.* 123, 2701–2716. doi: 10.1121/1.2902178

Kim, J., Matney, C. J., Blankenship, A., Hestrin, S., and Brown, S. P. (2014). Layer 6 Corticothalamic Neurons Activate a Cortical Output Layer, Layer 5a. *J. Neurosci.* 34, 9656–9664. doi: 10.1523/JNEUROSCI.1325-14.2014

Kimura, A., Donishi, T., Okamoto, K., and Tamai, Y. (2005). Topography of projections from the primary and non-primary auditory cortical areas to the medial geniculate body and thalamic reticular nucleus in the rat. *Neuroscience* 135, 1325–1342. doi: 10.1016/j.neuroscience.2005.06.089

Kimura, A., Donishi, T., Sakoda, T., Hazama, M., and Tamai, Y. (2003). Auditory thalamic nuclei projections to the temporal cortex in the rat. *Neuroscience* 117, 1003–1016. doi: 10.1016/s0306-4522(02)00949-1

Klatt, D. H. (1982). "Prediction of perceived phonetic distance from critical-band spectra: a first step," in *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, Paris, 1278–1281. doi: 10.1109/ICASSP.1982.1171512

Klepzig, K., Horn, U., König, J., Holtz, K., Wendt, J., Hamm, A. O., et al. (2020). Brain imaging of chill reactions to pleasant and unpleasant sounds. *Behav. Brain Res.* 380:112417. doi: 10.1016/j.bbr.2019.112417

Klinge, A., and Klump, G. M. (2009). Frequency difference limens of pure tones and harmonics within complex stimuli in Mongolian gerbils and humans. *J. Acoust. Soc. Am.* 125, 304–314. doi: 10.1121/1.3021315

Koelsch, S., Gunter, T. C., Cramon, D. Y., Zysset, S., Lohmann, G., and Friederici, A. D. (2002). Bach speaks: a cortical "language-network" serves the processing of music. *Neuroimage* 17, 956–966. doi: 10.1006/nimg.2002.1154

Kommajosyula, S. P., Bartlett, E. L., Cai, R., Ling, L., and Caspary, D. (2021). Corticothalamic projections deliver enhanced-responses to medial geniculate body as a function of the temporal reliability of the stimulus. *bioRxiv* [Preprint]. doi: 10.1101/2021.05.07.443156

Kommajosyula, S. P., Cai, R., Bartlett, E., and Caspary, D. M. (2019). Top-down or bottom up: decreased stimulus salience increases responses to predictable stimuli of auditory thalamic neurons. *J. Physiol.* 597, 2767–2784. doi: 10.1113/JP277450

Kraus, N., McGee, T. J., Carrell, T. D., Zecker, S. G., Nicol, T. G., and Koch, D. B. (1996). Auditory neurophysiologic responses and discrimination deficits in children with learning problems. *Science* 273, 971–973. doi: 10.1126/science.273.5277.971

Kraus, N., and Slater, J. (2015). Music and language: relations and disconnections. *Handb. Clin. Neurol.* 129, 207–222. doi: 10.1016/B978-0-444-62630-1.00012-3

Kraus, N., Slater, J., Thompson, E. C., Hornickel, J., Strait, D. L., Nicol, T., et al. (2014). Music enrichment programs improve the neural encoding of speech in at-risk children. *J. Neurosci.* 34, 11913–11918. doi: 10.1523/JNEUROSCI.1881-14.2014

Krause, V., Schnitzler, A., and Pollok, B. (2010). Functional network interactions during sensorimotor synchronization in musicians and non-musicians. *Neuroimage* 52, 245–251. doi: 10.1016/j.neuroimage.2010.03.081

Kudo, M., and Niimi, K. (1980). Ascending projections of the inferior colliculus in the cat: an autoradiographic study. *J. Comp. Neurol.* 191, 545–556. doi: 10.1002/cne.901910403

Lam, Y.-W., and Sherman, S. M. (2010). Functional organization of the somatosensory cortical layer 6 feedback to the thalamus. *Cereb. Cortex* 20, 13–24. doi: 10.1093/cercor/bhp077

Las, L., Stern, E. A., and Nelken, I. (2005). Representation of tone in fluctuating maskers in the ascending auditory system. *J. Neurosci.* 25, 1503–1513. doi: 10.1523/JNEUROSCI.4007-04.2005

LeDoux, J. E., Farb, C. R., and Romanski, L. M. (1991). Overlapping projections to the amygdala and striatum from auditory processing areas of the thalamus and cortex. *Neurosci. Lett.* 134, 139–144. doi: 10.1016/0304-3940(91)90526-Y

LeDoux, J. E., Ruggiero, D. A., and Reis, D. J. (1985). Projections to the subcortical forebrain from anatomically defined regions of the medial geniculate body in the rat. *J. Comp. Neurol.* 242, 182–213. doi: 10.1002/cne.902420204

Lee, C. C. (2013). Thalamic and cortical pathways supporting auditory processing. *Brain Lang.* 126, 22–28. doi: 10.1016/j.bandl.2012.05.004

Lee, C. C. (2015). Exploring functions for the non-lemniscal auditory thalamus. *Front. Neural Circuits* 9:69. doi: 10.3389/fncir.2015.00069

Lee, C. C., Lam, Y.-W., and Sherman, S. M. (2012). Intracortical convergence of layer 6 neurons. *Neuroreport* 23, 736–740. doi: 10.1097/WNR.0b013e328356c1aa

Lee, C.-C., and Middlebrooks, J. C. (2011). Auditory cortex spatial sensitivity sharpens during task performance. *Nat. Neurosci.* 14, 108–114. doi: 10.1038/nn.2713

Lee, C. C., and Sherman, S. M. (2009). Modulator property of the intrinsic cortical projection from layer 6 to layer 4. *Front. Syst. Neurosci.* 3:3. doi: 10.3389/neuro.06.003.2009

Lee, C. C., and Winer, J. A. (2008). Connections of cat auditory cortex: I. Thalamocortical system. *J. Comp. Neurol.* 507, 1879–1900. doi: 10.1002/cne.21611

Lee, K. M., Skoe, E., Kraus, N., and Ashley, R. (2009). Selective Subcortical Enhancement of Musical Intervals in Musicians. *J. Neurosci.* 29, 5832–5840. doi: 10.1523/JNEUROSCI.6133-08.2009

Leonard, M. K., and Chang, E. F. (2014). Dynamic speech representations in the human temporal lobe. *Trends Cogn. Sci.* 18, 472–479. doi: 10.1016/j.tics.2014.05.001

Li, L., and Ebner, F. F. (2007). Cortical Modulation of Spatial and Angular Tuning Maps in the Rat Thalamus. *J. Neurosci.* 27, 167–179. doi: 10.1523/JNEUROSCI.4165-06.2007

Licklider, J. C. R. (1951). A duplex theory of pitch perception. *Experientia* 7, 128–134. doi: 10.1007/BF02156143

Lin, J. Y., and Hartmann, W. M. (1998). The pitch of a mistuned harmonic: evidence for a template model. *J. Acoust. Soc. Am.* 103, 2608–2617. doi: 10.1121/1.422781

Llano, D. A., and Sherman, S. M. (2008). Evidence for nonreciprocal organization of the mouse auditory thalamocortical-corticothalamic projection systems. . *J. Comp. Neurol.* 507, 1209–1227. doi: 10.1002/cne.21602

Llinás, R. R., and Steriade, M. (2006). Bursting of Thalamic Neurons and States of Vigilance. *J. Neurophysiol.* 95, 3297–3308. doi: 10.1152/jn.00166.2006

Lohr, B., and Dooling, R. J. (1998). Detection of changes in timbre and harmonicity in complex sounds by zebra finches (*Taeniopygia guttata*) and budgerigars (*Melopsittacus undulatus*). *J. Comp. Psychol.* 112, 36–47. doi: 10.1037/0735-7036.112.1.36

Lohse, M., Bajo, V. M., and King, A. J. (2019). Development, organization and plasticity of auditory circuits: lessons from a cherished colleague. *Eur. J. Neurosci.* 49, 990–1004. doi: 10.1111/ejn.13979

Lohse, M., Bajo, V. M., King, A. J., and Willmore, B. D. B. (2020). Neural circuits underlying auditory contrast gain control and their perceptual implications. *Nat. Commun.* 11:324. doi: 10.1038/s41467-019-14163-5

Luo, F., Liu, X., Wang, C., and Yan, J. (2011). The pedunculopontine tegmental nucleus: a second cholinergic source for frequency-specific auditory plasticity. *J. Neurophysiol.* 105, 107–116. doi: 10.1152/jn.00546.2010

Luo, F., Wang, Q., Kashani, A., and Yan, J. (2008). Corticofugal modulation of initial sound processing in the brain. *J. Neurosci.* 28, 11615–11621. doi: 10.1523/JNEUROSCI.3972-08.2008

Ma, X., and Suga, N. (2007). Multiparametric corticofugal modulation of collicular duration-tuned neurons: modulation in the amplitude domain. *J. Neurophysiol.* 97, 3722–3730. doi: 10.1152/jn.01268.2006

Maddox, R. K., Atilgan, H., Bizley, J. K., and Lee, A. K. C. (2015). Auditory selective attention is enhanced by a task-irrelevant temporally coherent visual stimulus in human listeners. *eLife* 4:e04995. doi: 10.7554/eLife.04995

Massaux, A., Dutrieux, G., Cotillon-Williams, N., Manunta, Y., and Edeline, J.-M. (2004). Auditory thalamus bursts in anesthetized and non-anesthetized states: contribution to functional properties. *J. Neurophysiol.* 91, 2117–2134. doi: 10.1152/jn.00970.2003

McCormick, D. A., and Feeser, H. R. (1990). Functional implications of burst firing and single spike activity in lateral geniculate relay neurons. *Neuroscience* 39, 103–113. doi: 10.1016/0306-4522(90)90225-S

McDermott, J. H., Lehr, A. J., and Oxenham, A. J. (2010). Individual differences reveal the basis of consonance. *Curr. Biol.* 20, 1035–1041. doi: 10.1016/j.cub.2010.04.019

McDermott, J. H., and Oxenham, A. J. (2008). Music perception, pitch, and the auditory system. *Curr. Opin. Neurobiol.* 18, 452–463. doi: 10.1016/j.conb.2008.09.005

McDermott, J. H., Schultz, A. F., Undurraga, E. A., and Godoy, R. A. (2016). Indifference to dissonance in native Amazonians reveals cultural variation in music perception. *Nature* 535, 547–550. doi: 10.1038/nature18635

McDermott, J. H., and Simoncelli, E. P. (2011). Sound texture perception via statistics of the auditory periphery: evidence from sound synthesis. *Neuron* 71, 926–940. doi: 10.1016/j.neuron.2011.06.032

McPherson, M. J., and McDermott, J. H. (2020). Time-dependent discrimination advantages for harmonic sounds suggest efficient coding for memory. *Proc. Natl. Acad. Sci. U.S.A.*117, 32169–32180. doi: 10.1073/pnas.2008956117

Mease, R. A., Krieger, P., and Groh, A. (2014). Cortical control of adaptation and sensory relay mode in the thalamus. *Proc. Natl. Acad. Sci. U.S.A.* 111, 6798–6803. doi: 10.1073/pnas.1318665111

Meddis, R., and Hewitt, M. J. (1991a). Virtual pitch and phase sensitivity of a computer model of the auditory periphery. I: pitch identification. *J. Acoust. Soc. Am.* 89:2866. doi: 10.1121/1.400725

Meddis, R., and Hewitt, M. J. (1991b). Virtual pitch and phase sensitivity of a computer model of the auditory periphery. II: phase sensitivity. *J. Acoust. Soc. Am.* 89:2883. doi: 10.1121/1.400726

Menell, P., McAnally, K. I., and Stein, J. F. (1999). Psychophysical Sensitivity and Physiological Response to Amplitude Modulation in Adult Dyslexic Listeners. *J. Speech Lang. Hear. Res.* 42, 797–803. doi: 10.1044/jslhr.4204.797

Mesgarani, N., and Chang, E. F. (2012). Selective cortical representation of attended speaker in multi-talker speech perception. *Nature* 485, 233–236. doi: 10.1038/nature11020

Mesgarani, N., Cheung, C., Johnson, K., and Chang, E. F. (2014). Phonetic Feature Encoding in Human Superior Temporal Gyrus. *Science* 343, 1006–1010. doi: 10.1126/science.1245994

Micheyl, C., and Oxenham, A. J. (2010). Pitch, harmonicity and concurrent sound segregation: psychoacoustical and neurophysiological findings. *Hear. Res.* 266, 36–51. doi: 10.1016/j.heares.2009.09.012

Mihai, P. G., Moerel, M., de Martino, F., Trampel, R., Kiebel, S., and von Kriegstein, K. (2019). Modulation of tonotopic ventral medial geniculate body is behaviorally relevant for speech recognition. *Elife* 8:e44837. doi: 10.7554/eLife.44837

Mihai, P. G., Tschentscher, N., and von Kriegstein, K. (2021). Modulation of the primary auditory thalamus when recognising speech with background noise. *J. Neurosci.* doi: 10.1523/JNEUROSCI.2902-20.2021 [Epub ahead of print].

Miller, G. A., and Heise, G. A. (1950). The Trill Threshold. *J. Acoust. Soc. Am.* 22:637. doi: 10.1121/1.1906663

Miller, L. M., Escabí, M. A., Read, H. L., and Schreiner, C. E. (2002). Spectrotemporal receptive fields in the lemniscal auditory thalamus and cortex. *J. Neurophysiol.* 87, 516–527. doi: 10.1152/jn.00395.2001

Miller, L. M., and Schreiner, C. E. (2000). Stimulus-based state control in the thalamocortical system. *J. Neurosci.* 20, 7011–7016. doi: 10.1523/JNEUROSCI.20-18-07011.2000

Moerel, M., De Martino, F., Uğurbil, K., Yacoub, E., and Formisano, E. (2015). Processing of frequency and location in human subcortical auditory structures. *Sci. Rep.* 5:17048. doi: 10.1038/srep17048

Moore, B. C., Glasberg, B. R., and Peters, R. W. (1986). Thresholds for hearing mistuned partials as separate tones in harmonic complexes. *J. Acoust. Soc. Am.* 80, 479–483. doi: 10.1121/1.394043

Moore, B. C. J., Peters, R. W., and Glasberg, B. R. (1985). Thresholds for the detection of inharmonicity in complex tones. *J. Acoust. Soc. Am.* 77, 1861–1867. doi: 10.1121/1.391937

Morel, A., and Imig, T. J. (1987). Thalamic projections to fields A, AI, P, and VP in the cat auditory cortex. *J. Comp. Neurol.* 265, 119–144. doi: 10.1002/cne.902650109

Morel, A., Rouiller, E., de Ribaupierre, Y., and de Ribaupierre, F. (1987). Tonotopic organization in the medial geniculate body (MGB) of lightly anesthetized cats. *Exp. Brain Res.* 69, 24–42. doi: 10.1007/BF00247026

Müller, F., Niso, G., Samiee, S., Ptito, M., Baillet, S., and Kupers, R. (2019). A thalamocortical pathway for fast rerouting of tactile information to occipital cortex in congenital blindness. *Nat. Commun.* 10:5154. doi: 10.1038/s41467-019-13173-7

Musacchia, G., Large, E. W., and Schroeder, C. E. (2014). Thalamocortical mechanisms for integrating musical tone and rhythm. *Hear. Res.* 308, 50–59. doi: 10.1016/j.heares.2013.09.017

Nakamoto, K. T., Jones, S. J., and Palmer, A. R. (2008). Descending projections from auditory cortex modulate sensitivity in the midbrain to cues for spatial position. *J. Neurophysiol.* 99, 2347–2356. doi: 10.1152/jn.01326.2007

Nelson, S. L., Kong, L., Liu, X., and Yan, J. (2015). Auditory cortex directs the input-specific remodeling of thalamus. *Hear. Res.* 328, 1–7. doi: 10.1016/j.heares.2015.06.016

Niwa, M., Johnson, J. S., O'Connor, K. N., and Sutter, M. L. (2012). Active engagement improves primary auditory cortical neurons' ability to discriminate temporal modulation. *J. Neurosci.* 32, 9323–9334. doi: 10.1523/JNEUROSCI.5832-11.2012

Niwa, M., O'Connor, K. N., Engall, E., Johnson, J. S., and Sutter, M. L. (2015). Hierarchical effects of task engagement on amplitude modulation encoding in auditory cortex. *J. Neurophysiol.* 113, 307–327. doi: 10.1152/jn.00458.2013

Norman-Haignere, S., Kanwisher, N. G., and McDermott, J. H. (2015). Distinct cortical pathways for music and speech revealed by hypothesis-free voxel decomposition. *Neuron* 88, 1281–1296. doi: 10.1016/j.neuron.2015.11.035

Ohl, F. W., Wetzel, W., Wagner, T., Rech, A., and Scheich, H. (1999). Bilateral ablation of auditory cortex in Mongolian gerbil affects discrimination of frequency modulated tones but not of pure tones. *Learn. Mem.* 6, 347–362. doi: 10.1101/lm.6.4.347

Ojima, H. (1994). Terminal morphology and distribution of corticothalamic fibers originating from layers 5 and 6 of cat primary auditory cortex. *Cereb. Cortex* 4, 646–663. doi: 10.1093/cercor/4.6.646

Olsen, S. R., Bortone, D. S., Adesnik, H., and Scanziani, M. (2012). Gain control by layer six in cortical circuits of vision. *Nature* 483, 47–52. doi: 10.1038/nature10835

Ono, K., Kudoh, M., and Shibuki, K. (2006). Roles of the auditory cortex in discrimination learning by rats. *Eur. J. Neurosci.* 23, 1623–1632. doi: 10.1111/j.1460-9568.2006.04695.x

Osmanski, M. S., Song, X., and Wang, X. (2013). The role of harmonic resolvability in pitch perception in a vocal nonhuman primate, the common marmoset (*Callithrix jacchus*). *J. Neurosci.* 33, 9161–9168. doi: 10.1523/JNEUROSCI.0066-13.2013

O'Sullivan, J., Herrero, J., Smith, E., Schevon, C., McKhann, G. M., Sheth, S. A., et al. (2019). Hierarchical encoding of attended auditory objects in multi-talker speech perception. *Neuron* 104, 1195–1209.e3. doi: 10.1016/j.neuron.2019.09.007

Ozernov-Palchik, O., and Gaab, N. (2016). Tackling the "dyslexia paradox": reading brain and behavior for early markers of developmental dyslexia. *Wiley Interdiscip. Rev. Cogn. Sci.* 7, 156–176. doi: 10.1002/wcs.1383

Parbery-Clark, A., Skoe, E., and Kraus, N. (2009). Musical experience limits the degradative effects of background noise on the neural processing of sound. *J. Neurosci.* 29, 14100–14107. doi: 10.1523/JNEUROSCI.3256-09.2009

Parbery-Clark, A., Strait, D. L., and Kraus, N. (2011). Context-dependent encoding in the auditory brainstem subserves enhanced speech-in-noise perception in musicians. *Neuropsychologia* 49, 3338–3345. doi: 10.1016/j.neuropsychologia.2011.08.007

Patterson, R. D., Uppenkamp, S., Johnsrude, I. S., and Griffiths, T. D. (2002). The processing of temporal pitch and melody information in auditory cortex. *Neuron* 36, 767–776. doi: 10.1016/S0896-6273(02)01060-7

Peiffer, A. M., Rosen, G. D., and Fitch, R. H. (2002). Rapid auditory processing and MGN morphology in microgyric rats reared in varied acoustic environments. *Dev. Brain Res.* 138, 187–193. doi: 10.1016/S0165-3806(02)00472-8

Pellegrino, F., Coupe, C., and Marsico, E. (2011). A cross-language perspective on speech information rate. *Language (Baltim)* 87, 539–558. doi: 10.2307/23011654

Penagos, H., Melcher, J. R., and Oxenham, A. J. (2004). A neural representation of pitch salience in nonprimary human auditory cortex revealed with functional magnetic resonance imaging. *J. Neurosci.* 24, 6810–6815. doi: 10.1523/JNEUROSCI.0383-04.2004

Peruzzi, D., Bartlett, E., Smith, P. H., and Oliver, D. L. (1997). A monosynaptic GABAergic input from the inferior colliculus to the medial geniculate body in rat. *J. Neurosci.* 17, 3766–3777. doi: 10.1523/JNEUROSCI.17-10-03766.1997

Philibert, B., Laudanski, J., and Edeline, J. M. (2005). Auditory thalamus responses to guinea-pig vocalizations: a comparison between rat and guinea-pig. *Hear. Res.* 209, 97–103. doi: 10.1016/j.heares.2005.07.004

Pizzo, F., Roehri, N., Medina Villalon, S., Trébuchon, A., Chen, S., Lagarde, S., et al. (2019). Deep brain activities can be detected with magnetoencephalography. *Nat. Commun.* 10:971. doi: 10.1038/s41467-019-08665-5

Plomp, R., and Levelt, W. J. M. (1965). Tonal consonance and critical bandwidth. *J. Acoust. Soc. Am.* 38, 548–560. doi: 10.1121/1.1909741

Poeppel, D. (2003). The analysis of speech in different temporal integration windows: cerebral lateralization as 'asymmetric sampling in time.'. *Speech Commun.* 41, 245–255. doi: 10.1016/S0167-6393(02)00107-3

Polley, D. B., Read, H. L., Storace, D. A., and Merzenich, M. M. (2007). Multiparametric auditory receptive field organization across five cortical fields in the albino rat. *J. Neurophysiol.* 97, 3621–3638. doi: 10.1152/jn.01298.2006

Popham, S., Boebinger, D., Ellis, D. P. W., Kawahara, H., and McDermott, J. H. (2018). Inharmonic speech reveals the role of harmonicity in the cocktail party problem. *Nat. Commun.* 9:2122. doi: 10.1038/s41467-018-04551-8

Presacco, A., Simon, J. Z., and Anderson, S. (2016). Evidence of degraded representation of speech in noise, in the aging midbrain and cortex. *J. Neurophysiol.* 116, 2346–2355. doi: 10.1152/jn.00372.2016

Preuss, A., and Müller-Preuss, P. (1990). Processing of amplitude modulated sounds in the medial geniculate body of squirrel monkeys. *Exp. Brain Res.* 79, 207–211. doi: 10.1007/BF00228890

Quraishe, S., Newman, T., and Anderson, L. (2020). Auditory temporal acuity improves with age in the male mouse auditory thalamus: A role for perineuronal nets? *J. Neurosci. Res.* 98, 1780–1799. doi: 10.1002/jnr.24537

Rabinowitz, N. C., Willmore, B. D. B., Schnupp, J. W. H., and King, A. J. (2012). Spectrotemporal contrast kernels for neurons in primary auditory cortex. *J. Neurosci.* 32, 11271–11284. doi: 10.1523/JNEUROSCI.1715-12.2012

Read, H. L., Miller, L. M., Schreiner, C. E., and Winer, J. A. (2008). Two thalamic pathways to primary auditory cortex. *Neuroscience* 152, 151–159. doi: 10.1016/j.neuroscience.2007.11.026

Read, H. L., Nauen, D. W., Escabi, M. A., Miller, L. M., Schreiner, C. E., and Winer, J. A. (2011). Distinct core thalamocortical pathways to central and dorsal primary auditory cortex. *Hear. Res.* 274, 95–104. doi: 10.1016/j.heares.2010.11.010

Redies, H., Brandner, S., and Creutzfeldt, O. D. (1989). Anatomy of the auditory thalamocortical system of the guinea pig. *J. Comp. Neurol.* 282, 489–511. doi: 10.1002/cne.902820403

Richardson, B. D., Ling, L. L., Uteshev, V. V., and Caspary, D. M. (2013). Reduced GABAA receptor-mediated tonic inhibition in aged rat auditory thalamus. *J. Neurosci.* 33, 1218–1227. doi: 10.1523/JNEUROSCI.3277-12.2013

Roberts, B., and Bailey, P. J. (1993). Spectral pattern and the perceptual fusion of harmonics. II. A special status for added components?. *J. Acoust. Soc. Am.* 94, 3165–3177. doi: 10.1121/1.407222

Roberts, B., and Bailey, P. J. (1996). Spectral regularity as a factor distinct from harmonic relations in auditory grouping. *J. Exp. Psychol. Hum. Percept. Perform.* 22, 604–614. doi: 10.1037//0096-1523.22.3.604

Roberts, B., and Bregman, A. S. (1991). Effects of the pattern of spectral spacing on the perceptual fusion of harmonics. *J. Acoust. Soc. Am.* 90:3050. doi: 10.1121/1.401779

Roberts, B., and Brunstrom, J. M. (1998). Perceptual segregation and pitch shifts of mistuned components in harmonic complexes and in regular inharmonic complexes. *J. Acoust. Soc. Am.* 104, 2326–2338. doi: 10.1121/1.423771

Roberts, B., and Brunstrom, J. M. (2001). Perceptual fusion and fragmentation of complex tones made inharmonic by applying different degrees of frequency shift and spectral stretch. *J. Acoust. Soc. Am.* 110, 2479–2490. doi: 10.1121/1.1410965

Rodrigues-Dagaeff, C., Simm, G., De Ribaupierre, Y., Villa, A., De Ribaupierre, F., and Rouiller, E. M. (1989). Functional organization of the ventral division of the medial geniculate body of the cat: evidence for a rostro-caudal gradient of response properties and cortical projections. *Hear. Res.* 39, 103–125. doi: 10.1016/0378-5955(89)90085-3

Rosen, S. (1992). Temporal information in speech: acoustic, auditory and linguistic aspects. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 336, 367–373. doi: 10.1098/rstb.1992.0070

Rouiller, E., de Ribaupierre, Y., Toros-Morel, A., and de Ribaupierre, F. (1981). Neural coding of repetitive clicks in the medial geniculate body of cat. *Hear. Res.* 5, 81–100. doi: 10.1016/0378-5955(81)90028-9

Rouiller, E. M., Colomb, E., Capt, M., and de Ribaupierre, F. (1985). Projections of the reticular complex of the thalamus onto physiologically characterized regions of the medial geniculate body. *Neurosci. Lett.* 53, 227–232. doi: 10.1016/0304-3940(85)90190-9

Rouiller, E. M., and de Ribaupierre, F. (1985). Origin of afferents to physiologically defined regions of the medial geniculate body of the cat: ventral and dorsal divisions. *Hear. Res.* 19, 97–114. doi: 10.1016/0378-5955(85)90114-5

Rouiller, E. M., Rodrigues-Dagaeff, C., Simm, G., De Ribaupierre, Y., Villa, A., and De Ribaupierre, F. (1989). Functional organization of the medial division of the medial geniculate body of the cat: tonotopic organization, spatial distribution of response properties and cortical connections. *Hear. Res.* 39, 127–142. doi: 10.1016/0378-5955(89)90086-5

Rouiller, E. M., and Welker, E. (1991). Morphology of corticothalamic terminals arising from the auditory cortex of the rat: a Phaseolus vulgaris-leucoagglutinin (PHA-L) tracing study. *Hear. Res.* 56, 179–190. doi: 10.1016/0378-5955(91)90168-9

Rouiller, E. M., and Welker, E. (2000). A comparative analysis of the morphology of corticothalamic projections in mammals. *Brain Res. Bull.* 53, 727–741. doi: 10.1016/S0361-9230(00)00364-6

Ryugo, D. K., and Weinberger, N. M. (1976). Corticofugal modulation of the medial geniculate body. *Exp. Neurol.* 391, 377–391.

Saldeitis, K., Happel, M. F. K., Ohl, F. W., Scheich, H., and Budinger, E. (2014). Anatomy of the auditory thalamocortical system in the mongolian gerbil: nuclear origins and cortical field-, layer-, and frequency-specificities. *J. Comp. Neurol.* 522, 2397–2430. doi: 10.1002/cne.23540

Saldeitis, K., Jeschke, M., Budinger, E., Ohl, F. W., and Happel, M. F. K. (2021). Laser-induced apoptosis of corticothalamic neurons in layer VI of auditory cortex impact on cortical frequency processing. *Front. Neural Circuits* 15:659280. doi: 10.3389/fncir.2021.659280

Salvi, R. J., Lockwood, A. H., Frisina, R. D., Coad, M. L., Wack, D. S., and Frisina, D. R. (2002). PET imaging of the normal human auditory system: responses to speech in quiet and in background noise. *Hear. Res.* 170, 96–106. doi: 10.1016/S0378-5955(02)00386-6

Schall, S., Kiebel, S. J., Maess, B., and von Kriegstein, K. (2015). Voice identity recognition: functional division of the right STS and its behavioral relevance. *J. Cogn. Neurosci.* 27, 280–291. doi: 10.1162/jocn_a_00707

Schneider, D. M., and Woolley, S. M. N. (2013). Sparse and background-invariant coding of vocalizations in auditory scenes. *Neuron* 79, 141–152. doi: 10.1016/j.neuron.2013.04.038

Schön, D., Magne, C., and Besson, M. (2004). The music of speech: music training facilitates pitch processing in both music and language. *Psychophysiology* 41, 341–349. doi: 10.1111/1469-8986.00172.x

Schwartz, Z. P., and David, S. V. (2018). Focal suppression of distractor sounds by selective attention in auditory cortex. *Cereb. Cortex* 28, 323–339. doi: 10.1093/cercor/bhx288

Schwarz, D. W. F., and Tomlinson, R. W. W. (1990). Spectral response patterns of auditory cortex neurons to harmonic complex tones in alert monkey (*Macaca mulatta*). *J. Neurophysiol.* 64, 282–298. doi: 10.1152/jn.1990.64.1.282

Scott, B. H., Malone, B. J., and Semple, M. N. (2011). Transformation of temporal processing across auditory cortex of awake macaques. *J. Neurophysiol.* 105, 712–730. doi: 10.1152/jn.01120.2009

Scott, S. K. (2000). Identification of a pathway for intelligible speech in the left temporal lobe. *Brain* 123, 2400–2406. doi: 10.1093/brain/123.12.2400

Shackleton, T. M., and Carlyon, R. P. (1994). The role of resolved and unresolved harmonics in pitch perception and frequency modulation discrimination. *J. Acoust. Soc. Am.* 95, 3529–3540. doi: 10.1121/1.409970

Shamma, S. A., Elhilali, M., and Micheyl, C. (2011). Temporal coherence and attention in auditory scene analysis. *Trends Neurosci.* 34, 114–123. doi: 10.1016/j.tins.2010.11.002

Shamma, S., and Lorenzi, C. (2013). On the balance of envelope and temporal fine structure in the encoding of speech in the early auditory system. *J. Acoust. Soc. Am.* 133, 2818–2833. doi: 10.1121/1.4795783

Sherman, S. M. (2001). Tonic and burst firing: dual modes of thalamocortical relay. *Trends Neurosci.* 24, 122–126. doi: 10.1016/S0166-2236(00)01714-8

Sherman, S. M., and Guillery, R. W. (1998). On the actions that one nerve cell can have on another: distinguishing "drivers" from "modulators". *Proc. Natl. Acad. Sci. U.S.A.* 95, 7121–7126. doi: 10.1073/pnas.95.12.7121

Sherman, S. M., and Guillery, R. W. (2011). Distinct functions for direct and transthalamic corticocortical connections. *J. Neurophysiol.* 106, 1068–1077. doi: 10.1152/jn.00429.2011

Sherman, S. M., and Guillery, R. W. (2013). *Functional Connections of Cortical Areas: A New View from the Thalamus*, 1st Edn. Cambridge, MA: The MIT Press.

Sinex, D. G., Guzik, H., Li, H., Henderson Sabes, J., and Sabes, J. H. (2003). Responses of auditory nerve fibers to harmonic and mistuned complex tones. *Hear. Res.* 182, 130–139. doi: 10.1016/S0378-5955(03)00189-8

Sinex, D. G., Henderson Sabes, J., and Li, H. (2002). Responses of inferior colliculus neurons to harmonic and mistuned complex tones. *Hear. Res.* 168, 150–162. doi: 10.1016/S0378-5955(02)00366-0

Sinex, D. G., Li, H., and Velenovsky, D. S. (2005). Prevalence of stereotypical responses to mistuned complex tones in the inferior colliculus. *J. Neurophysiol.* 94, 3523–3537. doi: 10.1152/jn.01194.2004

Slater, J., Skoe, E., Strait, D. L., O'Connell, S., Thompson, E., and Kraus, N. (2015). Music training improves speech-in-noise perception: longitudinal evidence from a community-based music program. *Behav. Brain Res.* 291, 244–252. doi: 10.1016/j.bbr.2015.05.026

Smith, P. H., Uhlrich, D. J., Manning, K. A., and Banks, M. I. (2012). Thalamocortical projections to rat auditory cortex from the ventral and dorsal divisions of the medial geniculate nucleus. *J. Comp. Neurol.* 520, 34–51. doi: 10.1002/cne.22682

So, N. L. T., Edwards, J. A., and Woolley, S. M. N. (2020). Auditory selectivity for spectral contrast in cortical neurons and behavior. *J. Neurosci.* 40, 1015–1027. doi: 10.1523/JNEUROSCI.1200-19.2019

Souffi, S., Lorenzi, C., Varnet, L., Huetz, C., and Edeline, J.-M. (2020). Noise-sensitive but more precise subcortical representations coexist with robust cortical encoding of natural vocalizations. *J. Neurosci.* 40, 5228–5246. doi: 10.1523/JNEUROSCI.2731-19.2020

Souffi, S., Nodal, F. R., Bajo, V. M., and Edeline, J.-M. (2021). When and how does the auditory cortex influence subcortical auditory structures? New insights about the roles of descending cortical projections. *Front. Neurosci.* 15. doi: 10.3389/fnins.2021.690223

Stein, J. (2019). The current status of the magnocellular theory of developmental dyslexia. *Neuropsychologia* 130, 66–77. doi: 10.1016/j.neuropsychologia.2018.03.022

Stein, J. F., and McAnally, K. (1995). Auditory temporal processing in developmental dyslexics. *Irish J. Psychol.* 16, 220–228. doi: 10.1080/03033910.1995.10558058

Steinmetzger, K., and Rosen, S. (2015). The role of periodicity in perceiving speech in quiet and in background noise. *J. Acoust. Soc. Am.* 138, 3586–3599. doi: 10.1121/1.4936945

Steinschneider, M., Reser, D. H., Fishman, Y. I., Schroeder, C. E., and Arezzo, J. C. (1998). Click train encoding in primary auditory cortex of the awake monkey: evidence for two mechanisms subserving pitch perception. *J. Acoust. Soc. Am.* 104, 2935–2955. doi: 10.1121/1.423877

Steriade, M., McCormick, D., and Sejnowski, T. (1993). Thalamocortical oscillations in the sleeping and aroused brain. *Science* 262, 679–685. doi: 10.1126/science.8235588

Storace, D. A., Higgins, N. C., and Read, H. L. (2010). Thalamic label patterns suggest primary and ventral auditory fields are distinct core regions. *J. Comp. Neurol.* 518, 1630–1646. doi: 10.1002/cne.22345

Storace, D. A., Higgins, N. C., and Read, H. L. (2011). Thalamocortical pathway specialization for sound frequency resolution. *J. Comp. Neurol.* 519, 177–193. doi: 10.1002/cne.22501

Su, Y., and Delgutte, B. (2019). Pitch of harmonic complex tones: rate and temporal coding of envelope repetition rate in inferior colliculus of unanesthetized rabbits. *J. Neurophysiol.* 122, 2468–2485. doi: 10.1152/jn.00512.2019

Su, Y., and Delgutte, B. (2020). Robust rate-place coding of resolved components in harmonic and inharmonic complex tones in auditory midbrain. *J. Neurosci.* 40, 2080–2093. doi: 10.1523/JNEUROSCI.2337-19.2020

Sumner, C. J., Wells, T. T., Bergevin, C., Sollini, J., Kreft, H. A., Palmer, A. R., et al. (2018). Mammalian behavior and physiology converge to confirm sharper cochlear tuning in humans. *Proc. Natl. Acad. Sci. U.S.A.* 115, 11322–11326. doi: 10.1073/pnas.1810766115

Suta, D., Popelár, J., Kvasnák, E., and Syka, J. (2007). Representation of species-specific vocalizations in the medial geniculate body of the guinea pig. *Exp. Brain Res.* 183, 377–388. doi: 10.1007/s00221-007-1056-3

Swaminathan, J., Mason, C. R., Streeter, T. M., Best, V., Kidd, G. Jr., and Patel, A. D. (2015). Musical training, individual differences and the cocktail party problem. *Sci. Rep.* 5:11628. doi: 10.1038/srep11628

Swanepoel, R., Oosthuizen, D. J. J., and Hanekom, J. J. (2012). The relative importance of spectral cues for vowel recognition in severe noise. *J. Acoust. Soc. Am.* 132, 2652–2662. doi: 10.1121/1.4751543

Symmes, D., Alexander, G. E., and Newman, J. D. (1980). Neural processing of vocalizations and artificial stimuli in the medial geniculate body of squirrel monkey. *Hear. Res.* 3, 133–146. doi: 10.1016/0378-5955(80)90041-6

Tang, J., Yang, W., and Suga, N. (2012). Modulation of thalamic auditory neurons by the primary auditory cortex. *J. Neurophysiol.* 108, 935–942. doi: 10.1152/jn.00251.2012

Temereanca, S., and Simons, D. J. (2004). Functional topography of corticothalamic feedback enhances thalamic spatial response tuning in the somatosensory whisker/barrel system. *Neuron* 41, 639–651. doi: 10.1016/S0896-6273(04)00046-7

Terhardt, E. (1974). Pitch, consonance, and harmony. *J. Acoust. Soc. Am.* 55, 1061–1069. doi: 10.1121/1.1914648

Tervaniemi, M., Medvedev, S. V., Alho, K., Pakhomov, S. V., Roudas, M. S., van Zuijen, T. L., et al. (2000). Lateralized automatic auditory processing of phonetic versus musical information: a PET study. *Hum. Brain Mapp.* 10, 74–79.

Tervaniemi, M., Szameitat, A. J., Kruck, S., Schröger, E., Alter, K., De Baene, W., et al. (2006). From air oscillations to music and speech: functional magnetic resonance imaging evidence for fine-tuned neural networks in audition. *J. Neurosci.* 26, 8647–8652. doi: 10.1523/JNEUROSCI.0995-06.2006

Tomlinson, R. W. W., and Schwarz, D. W. F. (1988). Perception of the missing fundamental in nonhuman primates. *J. Acoust. Soc. Am.* 84, 560–565. doi: 10.1121/1.396833

Town, S. M., and Bizley, J. K. (2013). Neural and behavioral investigations into timbre perception. *Front. Syst. Neurosci.* 7:88. doi: 10.3389/fnsys.2013.00088

Tramo, M. J., De Cariani, P. A., Delgutte, B., and Braida, L. D. (2001). Neurobiological foundations for the theory of harmony in western tonal music. *Ann. N.Y. Acad. Sci.* 930, 92–116. doi: 10.1111/j.1749-6632.2001.tb05727.x

Tregellas, J. R., Ellis, J., Shatti, S., Du, Y. P., and Rojas, D. C. (2009). Increased hippocampal, thalamic, and prefrontal hemodynamic response to an urban noise stimulus in Schizophrenia. *Am. J. Psychiatry* 166, 354–360. doi: 10.1176/appi.ajp.2008.08030411

van Noorden, L. P. A. S. (1975). *Temporal Coherence in the Perception of Tone Sequences*. Ph.D. thesis. Eindhoven: Eindhoven University of Technology.

Vaquero, L., Hartmann, K., Ripollés, P., Rojo, N., Sierpowska, J., François, C., et al. (2016). Structural neuroplasticity in expert pianists depends on the age of musical training onset. *Neuroimage* 126, 106–119. doi: 10.1016/j.neuroimage.2015.11.008

Vasquez-Lopez, S. A., Weissenberger, Y., Lohse, M., Keating, P., King, A. J., and Dahmen, J. C. (2017). Thalamic input to auditory cortex is locally heterogeneous but globally tonotopic. *eLife* 6:e25141. doi: 10.7554/eLife.25141

Velenovsky, D. S., Cetas, J. S., Price, R. O., Sinex, D. G., and McMullen, N. T. (2003). Functional subregions in primary auditory cortex defined by thalamocortical terminal arbors: an electrophysiological and anterograde labeling study. *J. Neurosci.* 23, 308–316.

Villa, A. E. P. (1990). Physiological differentiation within the auditory part of the thalamic reticular nucleus of the cat. *Brain Res. Rev.* 15, 25–40. doi: 10.1016/0165-0173(90)90010-L

Villa, A. E. P., Rouiller, E. M., Simm, G. M., Zurita, P., de Ribaupierre, Y., and de Ribaupierre, F. (1991). Corticofugal modulation of the information processing in the auditory thalamus of the cat. *Exp. Brain Res.* 86, 506–517. doi: 10.1007/BF00230524

von Kriegstein, K., Patterson, R. D., and Griffiths, T. D. (2008). Task-dependent modulation of medial geniculate body is behaviorally relevant for speech recognition. *Curr. Biol.* 18, 1855–1859. doi: 10.1016/j.cub.2008.10.052

Walker, K. M. M., Gonzalez, R., Kang, J. Z., McDermott, J. H., and King, A. J. (2019). Across-species differences in pitch perception are consistent with differences in cochlear filtering. *eLife* 8:e41626. doi: 10.7554/eLife.41626

Walker, K. M. M., Schnupp, J. W. H., Hart-Schnupp, S. M. B., King, A. J., and Bizley, J. K. (2009). Pitch discrimination by ferrets for simple and complex sounds. *J. Acoust. Soc. Am.* 126, 1321–1335. doi: 10.1121/1.3179676

Wallace, M. N., Anderson, L. A., and Palmer, A. R. (2007). Phase-locked responses to pure tones in the auditory thalamus. *J. Neurophysiol.* 98, 1941–1952. doi: 10.1152/jn.00697.2007

Wang, W., Andolina, I. M., Lu, Y., Jones, H. E., and Sillito, A. M. (2016). Focal gain control of thalamic visual receptive fields by layer 6 corticothalamic feedback. *Cereb. Cortex* 28, 267–280. doi: 10.1093/cercor/bhw376

Weinberger, N. M. (2011). The medial geniculate, not the amygdala, as the root of auditory fear conditioning. *Hear. Res.* 274, 61–74. doi: 10.1016/j.heares.2010.03.093

Weyand, T. G., Boudreaux, M., and Guido, W. (2001). Burst and tonic response modes in thalamic neurons during sleep and wakefulness. *J. Neurophysiol.* 85, 1107–1118. doi: 10.1152/jn.2001.85.3.1107

Williamson, R. S., and Polley, D. B. (2019). Parallel pathways for sound processing and functional connectivity among layer 5 and 6 auditory corticofugal neurons. *eLife* 8:e42974. doi: 10.7554/eLife.42974

Winer, J. A. (2006). Decoding the auditory corticofugal systems. *Hear. Res.* 212, 1–8. doi: 10.1016/j.heares.2005.06.014

Winer, J. A., Diehl, J. J., and Larue, D. T. (2001). Projections of auditory cortex to the medial geniculate body of the cat. *J. Comp. Neurol.* 430, 27–55.

Winer, J. A., and Larue, D. T. (1987). Patterns of reciprocity in auditory thalamocortical and corticothalamic connections: study with horseradish peroxidase and autoradiographic methods in the rat medial geniculate body. *J. Comp. Neurol.* 257, 282–315. doi: 10.1002/cne.902570212

Winer, J. A., and Larue, D. T. (1996). Evolution of GABAergic circuitry in the mammalian medial geniculate body. *Proc. Natl. Acad. Sci. U.S.A.* 93, 3083–3087. doi: 10.1073/pnas.93.7.3083

Winer, J. A., and Prieto, J. J. (2001). Layer V in cat primary auditory cortex (AI): cellular architecture and identification of projection neurons. *J. Comp. Neurol.* 434, 379–412. doi: 10.1002/cne.1183

Winer, J. A., Saint Marie, R. L., Larue, D. T., and Oliver, D. L. (1996). GABAergic feedforward projections from the inferior colliculus to the medial geniculate body. *Proc. Natl. Acad. Sci. U.S.A.* 93, 8005–8010. doi: 10.1073/pnas.93.15.8005

Wong, P. C. M., Skoe, E., Russo, N. M., Dees, T., and Kraus, N. (2007). Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nat. Neurosci.* 10, 420–422. doi: 10.1038/nn1872

Xie, F., You, L., Cai, D., Liu, M., Yue, Y., Wang, Y., et al. (2018). Fast inhibitory decay facilitates adult-like temporal processing in layer 5 of developing primary auditory cortex. *Cereb. Cortex* 28, 4319–4335. doi: 10.1093/cercor/bhx284

Yan, J., and Ehret, G. (2002). Corticofugal modulation of midbrain sound processing in the house mouse. *Eur. J. Neurosci.* 16, 119–128. doi: 10.1046/j.1460-9568.2002.02046.x

Yin, P., Johnson, J. S., O'Connor, K. N., and Sutter, M. L. (2011). Coding of amplitude modulation in primary auditory cortex. *J. Neurophysiol.* 105, 582–600. doi: 10.1152/jn.00621.2010

Yu, Y.-Q., Xiong, Y., Chan, Y.-S., and He, J. (2004). Corticofugal gating of auditory information in the thalamus: an *in vivo* intracellular recording study. *J. Neurosci.* 24, 3060–3069. doi: 10.1523/JNEUROSCI.4897-03.2004

Zatorre, R., Evans, A., and Meyer, E. (1994). Neural mechanisms underlying melodic perception and memory for pitch. *J. Neurosci.* 14, 1908–1919. doi: 10.1523/JNEUROSCI.14-04-01908.1994

Zatorre, R. J., Belin, P., and Penhune, V. B. (2002). Structure and function of auditory cortex: music and speech. *Trends Cogn. Sci.* 6, 37–46. doi: 10.1016/S1364-6613(00)01816-7

Zatorre, R. J., Halpern, A. R., Perry, D. W., Meyer, E., and Evans, A. C. (1996). Hearing in the Mind's Ear: a PET investigation of musical imagery and perception. *J. Cogn. Neurosci.* 8, 29–46. doi: 10.1162/jocn.1996.8.1.29

Zatorre, R. J., and Schönwiesner, M. (2011). "Cortical speech and music processes revealed by functional neuroimaging," in *The Auditory Cortex*, eds J. A. Winer and C.E. Schreiner (Berlin: Springer), 657–677. doi: 10.1007/978-1-4419-0074-6_31

Zendel, B. R., and Alain, C. (2012). Musicians experience less age-related decline in central auditory processing. *Psychol. Aging* 27, 410–417. doi: 10.1037/a0024816

Zendel, B. R., Tremblay, C.-D., Belleville, S., and Peretz, I. (2015). The impact of musicianship on the cortical mechanisms related to separating speech from background noise. *J. Cogn. Neurosci.* 27, 1044–1059. doi: 10.1162/jocn_a_00758

Zhan, X. J., Cox, C. L., Rinzel, J., and Sherman, S. M. (1999). Current clamp and modeling studies of low-threshold calcium spikes in cells of the Cat's Lateral Geniculate Nucleus. *J. Neurophysiol.* 81, 2360–2373. doi: 10.1152/jn.1999.81.5.2360

Zhang, Y., and Suga, N. (2000). Modulation of responses and frequency tuning of thalamic and collicular neurons by cortical activation in mustached bats. *J. Neurophysiol.* 84, 325–333. doi: 10.1152/jn.2000.84.1.325

Zhang, Y., Suga, N., and Yan, J. (1997). Corticofugal modulation of frequency processing in bat auditory system. *Nature* 387, 900–903. doi: 10.1038/43180

Zhang, Y., Wang, X., Zhu, L., Bai, S., Li, R., Sun, H., et al. (2021). Selective corticofugal modulation on sound processing in auditory thalamus of awake marmosets. *bioRxiv* [Preprint]. doi: 10.1101/2021.03.13.435231

Zhang, Y., and Yan, J. (2008). Corticothalamic feedback for sound- specific plasticity of auditory thalamic neurons elicited by tones paired with basal forebrain stimulation. *Cereb. Cortex* 18, 1521–1528. doi: 10.1093/cercor/bhm188

Zhang, Z., Liu, C.-H., Yu, Y.-Q., Fujimoto, K., Chan, Y.-S., and He, J. (2008). Corticofugal projection inhibits the auditory thalamus through the thalamic reticular nucleus. *J. Neurophysiol.* 99, 2938–2945. doi: 10.1152/jn.00002.2008

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Homma and Bajo. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 03 September 2021
doi: 10.3389/fnins.2021.704805

Olivocochlear Changes Associated With Aging Predominantly Affect the Medial Olivocochlear System

*Sergio Vicencio-Jimenez[1], Madison M. Weinberg[1], Giuliana Bucci-Mansilla[2] and Amanda M. Lauer[1]**

[1] The Center for Hearing and Balance, Department of Otolaryngology-Head and Neck Surgery, Johns Hopkins University School of Medicine, Baltimore, MD, United States, [2] Laboratorio de Neurosistemas, Departamento de Neurociencia, Facultad de Medicina, Universidad de Chile, Santiago, Chile

Age-related hearing loss (ARHL) is a public health problem that has been associated with negative health outcomes ranging from increased frailty to an elevated risk of developing dementia. Significant gaps remain in our knowledge of the underlying central neural mechanisms, especially those related to the efferent auditory pathways. Thus, the aim of this study was to quantify and compare age-related alterations in the cholinergic olivocochlear efferent auditory neurons. We assessed, in young-adult and aged CBA mice, the number of cholinergic olivocochlear neurons, auditory brainstem response (ABR) thresholds in silence and in presence of background noise, and the expression of excitatory and inhibitory proteins in the ventral nucleus of the trapezoid body (VNTB) and in the lateral superior olive (LSO). In association with aging, we found a significant decrease in the number of medial olivocochlear (MOC) cholinergic neurons together with changes in the ratio of excitatory and inhibitory proteins in the VNTB. Furthermore, in old mice we identified a correlation between the number of MOC neurons and ABR thresholds in the presence of background noise. In contrast, the alterations observed in the lateral olivocochlear (LOC) system were less significant. The decrease in the number of LOC cells associated with aging was 2.7-fold lower than in MOC and in the absence of changes in the expression of excitatory and inhibitory proteins in the LSO. These differences suggest that aging alters the medial and lateral olivocochlear efferent pathways in a differential manner and that the changes observed may account for some of the symptoms seen in ARHL.

Keywords: superior olivary complex, auditory efferents, olivocochlear system, age related hearing loss, aging

OPEN ACCESS

Edited by:
Erika Skoe,
University of Connecticut, United States

Reviewed by:
Ruili Xie,
The Ohio State University, United States
Douglas E. Vetter,
University of Mississippi Medical Center, United States
Dwayne D. Simmons,
Baylor University, United States

***Correspondence:**
Amanda M. Lauer
alauer2@jhmi.edu

Specialty section:
This article was submitted to Auditory Cognitive Neuroscience, a section of the journal Frontiers in Neuroscience

Received: *03 May 2021*
Accepted: *02 August 2021*
Published: *03 September 2021*

Citation:
Vicencio-Jimenez S, Weinberg MM, Bucci-Mansilla G and Lauer AM (2021) Olivocochlear Changes Associated With Aging Predominantly Affect the Medial Olivocochlear System. Front. Neurosci. 15:704805. doi: 10.3389/fnins.2021.704805

INTRODUCTION

As social animals, any circumstance that disrupts our ability to communicate can have profound consequences on our health. Age-related hearing loss (ARHL) is the most common sensory impairment among the elderly and is the third leading health condition overall in older adults (Collins, 1997). ARHL is defined as a progressive loss of hearing ability that is most pronounced at high frequencies (Bowl and Dawson, 2019). Notably, it is also characterized by the deterioration of sound localization ability and a reduction in speech recognition, especially in noisy environments

(Dubno et al., 1984; Stuart and Phillips, 1996; Frisina and Frisina, 1997; Gordon-Salant, 2005; Shojaei et al., 2016). About half of people over 70 years of age have hearing impairments severe enough to reduce their communication abilities (Agrawal et al., 2008; Yamasoba et al., 2013), which is associated with social isolation, depression, accelerated cognitive decline, and increased risk of dementia (Lin, 2011; Livingston et al., 2017; Loughrey et al., 2018; Rutherford et al., 2018). Despite its prevalence and the negative health outcomes associated with ARHL, our understanding of the biological mechanisms and processes that explain this condition is still incomplete, particularly with respect to our knowledge of how the central auditory pathways are altered and the role they play in aging. These gaps in our knowledge of ARHL are even more prominent with respect to the efferent auditory pathways, which are not usually the focus of aging research.

In mammals, the auditory efferent pathways form a network composed of feedback loops that includes the auditory cortex and subcortical nuclei such as the thalamus, inferior colliculus, superior olivary complex and cochlear nucleus (Malmierca and Ryugo, 2011). Although recent results have presented evidence of direct projections to the cochlea from the ventral nucleus of the lateral lemniscus (Suthakar and Ryugo, 2021), virtually all efferent projections from the central nervous system to the cochlea leave from the superior olive. Thus, the final component of this efferent system that reaches the cochlea is known as the olivocochlear (OC) system, which originates in the superior olivary complex (SOC) (Rasmussen, 1946). The OC system is comprised of two neuronal groups: (i) the medial olivocochlear neurons (MOC) and (ii) the lateral olivocochlear neurons (LOC) (Warr and Guinan, 1979). Large MOC neurons are located in the medial periolivary region, predominantly in the ventral nucleus of the trapezoid body (VNTB) in rodents and send myelinated projections that make axo-somatic synapses with the outer hair cells of the cochlea (OHC) (Guinan et al., 1983; Brown, 2011; Fuchs and Lauer, 2018). Synapses are organized tonotopically, with greater density in the middle regions of the cochlea in most species (Guinan, 1996; Maison et al., 2003). LOC neurons are smaller than their MOC counterparts and originate in the lateral superior olive (LSO) (Guinan et al., 1983; Brown, 2011). They send unmyelinated fibers to the cochlea that make axo-axonal synapses with auditory nerve type I fibers near the inner hair cells (Guinan, 1996; Simmons, 2002). They also display a tonotopic organization, with slightly more innervation in the apical half of the cochlea (Guinan et al., 1984; Robertson et al., 1987; Liberman et al., 1990). Both MOC and LOC neuron populations release acetylcholine as their main synaptic transmitter (Bobbin and Konishi, 1971; Vetter et al., 1991; Eybalin, 1993; Blanchet et al., 1996). LOC neurons express a greater diversity of neurotransmitters than MOCs, including dopamine (DA), calcitonin gene-related peptide (CGRP), GABA, and opioid peptides such as enkephalin (Eybalin, 1993; Ciuman, 2010; Reijntjes and Pyott, 2016; Wu et al., 2020).

Activation of the MOC pathway produces a reduction in cochlear sensitivity (Mountain, 1980; Murugasu and Russell, 1996; Cooper and Guinan, 2006). Functionally, the MOC pathway has been shown to facilitate stimulus discrimination in noise, enhance auditory selective attention, and protect against damage from noise exposure (Winslow and Sachs, 1987; Kawase and Liberman, 1993; May et al., 2004; Terreros et al., 2016; Lauer et al., 2021). Less is known about the function of the LOC system, with current evidence being limited and sometimes contradictory. However, evidence suggests that LOC neurons can modulate the activity of type I auditory nerve fibers (Felix and Ehrenberger, 1992; Groff and Liberman, 2003; Wu et al., 2020). The efferent activity of the LOC potentially protects IHC and nerve fibers from acoustic overexposure (Wu et al., 2020) or could be modulating the set point of auditory nerve fibers, thus contributing to the generation a range of spontaneous firing rates (Le Prell et al., 2003; Ciuman, 2010; Wu et al., 2020).

Most of the available evidence for the role of OC pathways in ARHL comes from physiological assessments of the OC system and studies of inner ear structures. Rodent models have shown that age-related changes in the MOC system plays a role in the progression of ARHL (Jacobson et al., 2003; Zettel et al., 2007; Zhu et al., 2007; Boero et al., 2020). Also, experimental lesions of the olivocochlear bundle suggest that efferent feedback contributes to slowing cochlear aging (Liberman et al., 2014). In addition to this, there is evidence that OC pathways are weakened in aging humans (Parthasarathy, 2001; Kim et al., 2002). Furthermore, changes in cochlear efferent innervation have been observed in aging mice and humans (Lauer et al., 2012; Zachary and Fuchs, 2015; Liberman and Liberman, 2019; Jeng et al., 2020; Kobrina et al., 2020).

There is remarkably little work focused on alterations at the central level of the aging olivocochlear system (Radtke-Schuller et al., 2015). In the present work, we studied age-associated changes in the brainstem regions of the olivocochlear auditory efferent system of mice. We quantified the number of cholinergic OC neurons at different ages and compared them with auditory brainstem response (ABR) measurements in quiet and noise conditions. In addition, markers of excitatory [vesicular glutamate transporter 1 (VGLUT1)] and inhibitory [glutamic acid decarboxylase 65-kilodalton isoform (GAD65)] synapses were also assessed. These results were compared with age-related cell loss in the vestibular efferent and trigeminal motor nucleus to determine if age-related degeneration is specific to the auditory efferents.

MATERIALS AND METHODS

Animals

The experiments were performed using a total of 65 adult CBA/CaJ mice (30 males and 35 females). The original breeding pairs were obtained from The Jackson Laboratory (strain #000654). Mice were bred and housed in a quiet, low-traffic vivarium at Johns Hopkins University (Wu et al., 2020). The mice were housed in groups with *ad libitum* water and food and under a 12–12-h night/day cycle. This strain was selected because CBA/CaJ mice show a hearing loss and cochlear damage trajectory across the lifespan that is similar to humans (Sergeyenko et al., 2013; Kobrina and Dent, 2020; Kobrina et al., 2020). The subjects ranged from 1 to 30 months

old. Mice in the range of 1–8 months (n = 33; 18 males and 15 females) were considered young adults and those in the range of 18–30 months were considered old (n = 32; 12 males and 20 females). All procedures were approved by the Johns Hopkins University Institutional Animal Care and Use Committee and followed NIH guidelines for the care and use of laboratory animals.

Auditory Brainstem Response (ABR)

To evaluate hearing status, ABRs were recorded in quiet and noise backgrounds. Mice were anesthetized with an i.p. injection of 100 mg/kg ketamine and 20 mg/kg xylazine. To prevent corneal damage, an ophthalmic ointment was applied to the eyes. Once anesthetized, mice were placed inside a sound-attenuating chamber 10 cm from a speaker (MF1, TDT), measured from the pinnae. Mice were placed on a heating pad to maintain a temperature of 37°C ± 1°C. Subdermal platinum needle electrodes were placed at the vertex of the skull (+electrode), over the left bulla (−electrode) and in the muscle of the ipsilateral hind leg (ground electrode).

The ABR signal recorded by the electrodes was sent to a head stage (Medusa4Z, TDT) and then transferred to a pre-amplifier (RA4PA, TDT) with 20-fold amplification. The signals were acquired with a sampling rate of 12 kHz and bandpass filtered (highpass 0.3 kHz and lowpass 5 kHz), notched at 50 Hz, and averaged over 512 presentations. Off-line and prior to analysis, the signal was filtered again at 0.3–3 KHz.

The SigGenRZ software (TDT) was used to program the stimulus protocols for click and tones and generated using the TDT BioSigRZ platform. The stimuli consisted of clicks (0.1 ms square wave pulse of alternating polarity) and 5 ms tone pulses of 8, 12, 16, and 24 kHz (0.5 ms onset/offset), played at a rate of 21 repetitions/s. We decided not to evaluate the 32 KHz frequency because, in our experience, it was unlikely to find an ABR response at this frequency in old animals (Kobrina et al., 2020). All stimuli were presented at descending levels in 10 dB increments, starting at 90 to 0 dB. Stimuli were calibrated using a 1/4″ free-field microphone (Bruel & Kjaer) placed at the location where the mouse head would normally be during testing, using the BioSigRZ software (TDT). For the experiment in which the ABR stimuli were masked with background noise, a second speaker (MF1, TDT) was used, which was positioned 15 cm from the animal's right ear and at 90° with respect to the stimulus presentation (ABR) speaker. Through this speaker, and during the entire presentation of the ABR stimuli, a broadband noise of 40 dB SPL was presented. This intensity was measured at the level of the right pinna of the mouse.

Testing lasted between 70 and 90 min. After the recording was finished mice were placed in an individual cage over a heating pad and monitored until recovered. Once fully awake, they were returned to their home cages.

Auditory brainstem response recordings were analyzed offline using BioSigRZ and MATLAB (vR2019a; MathWorks). The amplitudes and latencies of waves 1 to 3 (click responses at 90 dB SPL) were measured offline manually by two independent examiners blinded to the animals' condition. Wave 4 and 5 were not reliably detected in the recording with background noise, so we did not quantify them. The ABR threshold was calculated automatically, using a custom MATLAB script. The threshold was defined as the sound intensity level at which the peak-to-trough amplitude of the ABR wave was at least two standard deviations above the mean baseline amplitude estimated from the last 5 ms in the recording when no sound stimulus was present (Lina and Lauer, 2013; McGuire et al., 2015; Lauer, 2017; Schrode et al., 2018). A value of 95 dB SPL was used as a threshold for cases where one could not be found.

Brainstem Sectioning

Mice were deeply anesthetized with sodium pentobarbital (i.p., 150 mg/kg) followed by an intracardiac injection of heparin (0.7 ml/kg). Immediately, the mice were transcardially perfused with 60 ml of a 4% paraformaldehyde solution. The brain tissue was removed and postfixed in 4% paraformaldehyde at 4°C for a minimum of three and up to 18 h. To perform the brain sections, the frontal portion of the cerebrum was removed (1 mm rostral from the confluence of the sinuses) allowing the regions of interest (ROI) to sit in their coronal plane. This remaining portion of the brain was embedded in a solution containing 5 ml of albumin gel mixed with 0.4 ml of 5% glutaraldehyde and 1 ml of 37% paraformaldehyde. The brain was mounted on a LeicaVT1200S vibratome and sectioned in the coronal plane in 50 μm slices. Serial sections through the auditory brainstem were placed in Tris-NaCl buffer (TBS).

Immunostaining

The brain sections were first incubated in a permeabilizing solution (TBS with 0.5% Triton X-100) for 15 min and blocking buffer (TBS with 10% normal rabbit serum) for 1 hr. The sections were incubated with the primary antibody for choline acetyltransferase (CHAT) diluted in TBS (1:400; Millipore AB144P) and 0.5% of Triton X-100 for 20 h at room temperature on a shaker. Afterward, the brain sections were incubated with secondary antibodies diluted in TBS (1:200 Biotinylated Rabbit Anti-Goat; Vector Labs BA-5000), for 1 h at room temperature on a shaker. Then, they were incubated with ABC reagent for 1 h at room temperature (ABC kit PK-6100) on a shaker. Finally, sections were treated for 5 ± 1 min with a solution containing 0.05% DAB, 0.4% nickel ammonium sulfate, and 0.01% hydrogen peroxide. In order to dilute and wash the reagents, the sections were rinsed with TBS three times between each of the steps described above.

In the case of the GAD65 and the VGLUT1 antibodies the procedures were similar, except for the primary (anti-GAD65 1:1000; Abcam ab26113 and anti-VGLUT1 1:1000; Thermo Fisher Scientific 48-2400) and secondary antibodies (1:200 Biotinylated goat anti-mouse; Vector Labs BA-9200 and 1:200 Biotinylated goat anti-rabbit; Vector Labs BA-1000 for GAD65 and VGLUT1, respectively). For more details, see Schrode et al. (2018).

To have negative controls, in each brain we omitted the primary antibody in one section and the secondary antibody in an additional section. Finally, the stained sections were mounted on glass slides, air-dried, mounted with Permount (Fisher Scientific), and coverslipped.

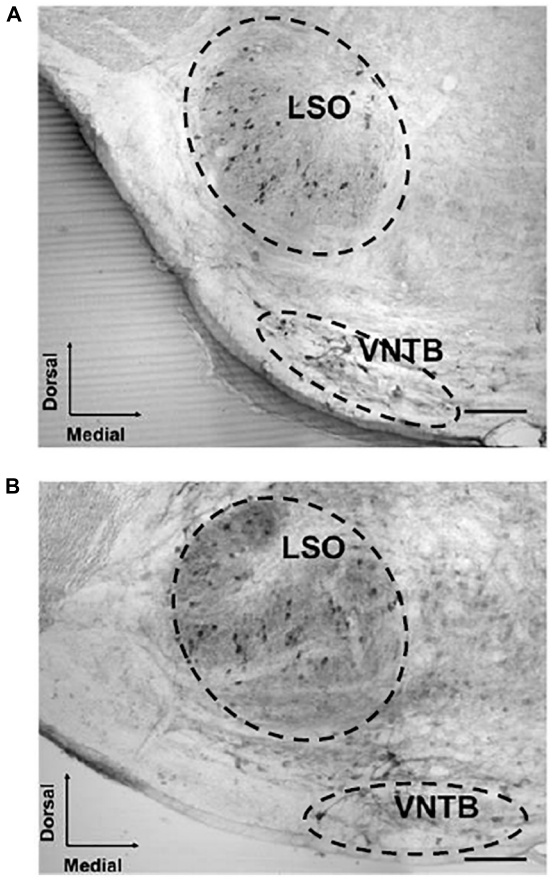

FIGURE 1 | Examples of acetylcholinesterase-stained sections of the SOC of a young and aged mouse. The image shows two photomicrographs (20×) of a cross section through the left side of the mouse brainstem. The image in panel **(A)** corresponds to a 2-month-old male mouse and panel **(B)** to a 26-month-old male mouse section. The somas of MOC neurons are located in the VNTB and LOC neurons can be seen in the LSO. 100 μm scale bar. LOC, lateral olivocochlear system; LSO, lateral superior olive; MOC, medial olivocochlear system; SOC, superior olivary complex; VNTB, ventral nucleus of the trapezoid body.

Cell and Area Quantification in Regions of Interest

Sections including the LSO, VNTB, vestibular efferent nuclei, and trigeminal motor nuclei were photographed at 40× on a microscope (Labophot; Nikon) with a mounted CCD camera (Progres; Jenoptik). To include both the LSO and VNTB in the same image, overlapping photographs were taken and merged using Fiji (Rueden et al., 2017). This procedure was also performed to combine the images of the trigeminal nuclei. The ROI were identified manually using a graphics tablet and stylus (Cintiq 22HD; Wacom). An automatic threshold algorithm was used in Fiji to identify the immunoreactive areas for CHAT, VGLUT1, and GAD65 antibodies. To determine the best suitable algorithm, all available automatic threshold algorithms in Fiji were evaluated against manual results from a blind observer in a reduced data set. It was determined that of the algorithms tested, Default, RenjiEntropy and Triangle algorithm was the best at identifying immunoreactive zones for CHAT, VGLUT1, and GAD65, respectively. To correct for illumination or staining intensity irregularities, we applied a leveling adjustment to each section, so that the mean pixel intensity within each ROI was equivalent across all sections of each animal (Schrode et al., 2018). To identify positive immunolabeling, we made a histogram of pixel intensity based on all pixels within the ROIs across all slices (Schrode et al., 2018). Label density was quantified as the fraction between the total number of thresholded pixels and the total number of pixels within each ROI.

In each of the sections, the number of positive CHAT cells per region of interest was determined using the Fiji cell counting plugin. The cell numbers shown correspond to the values of each ROI from one hemisphere. Cell counting was done by an observer who was blind to the age and hearing status of the animals. To reduce the possibility of double counting, cell count estimates were corrected using the Abercrombie method (Abercrombie, 1946). A second independent observer counted a fraction of the samples (25% of the total). No significant differences were found between the two observers.

Statistical Analysis

For our comparisons between young and old mice we determined that animals between 2 and 8 months of age qualified as part of the young adult group, while those older than 18 months qualified as old mice. This criterion was based on previous aging brackets determined in mice (Flurkey et al., 2007).

Statistical analyses were performed with MATLAB or GraphPad prism. Statistical test information and sample numbers are specified in the Figure Legends and in the Results. The error bars correspond to the standard error of the mean (SEM). For comparisons between two different groups (such as the number of cells between young and old mice) we used an unpaired *t* test with Welch's correction. For cases in which we evaluated how a response (such as ABR threshold) was affected by two factors (age and presence of background noise) we used a two-way ANOVA or a mixed effect analysis, both with a Sidak's multiple comparisons test. Statistical significance was defined as: $p > 0.05$ not significant (n.s.), $p < 0.05$ (*), $p < 0.001$ (**), and $p < 0.0001$ (***).

RESULTS

Olivocochlear Cell Count

We assessed the number of CHAT-immunoreactive cells in the SOC in 23 young mice (1–8 months of age) and in 20 old mice (18–30 months of age). In the SOC we identified two large populations of neurons that had dark staining, one located in the VNTB and the other in the LSO (**Figure 1**). In the VNTB we found larger cells that we classified as neurons of the MOC system, whereas those found in the LSO were classified as part of the LOC system (Brown and Levine, 2008; Radtke-Schuller et al., 2015). The number of total OC cells and their respective MOC and LOC subdivisions for the different ages are shown in **Figure 2**. We evaluated potential differences in

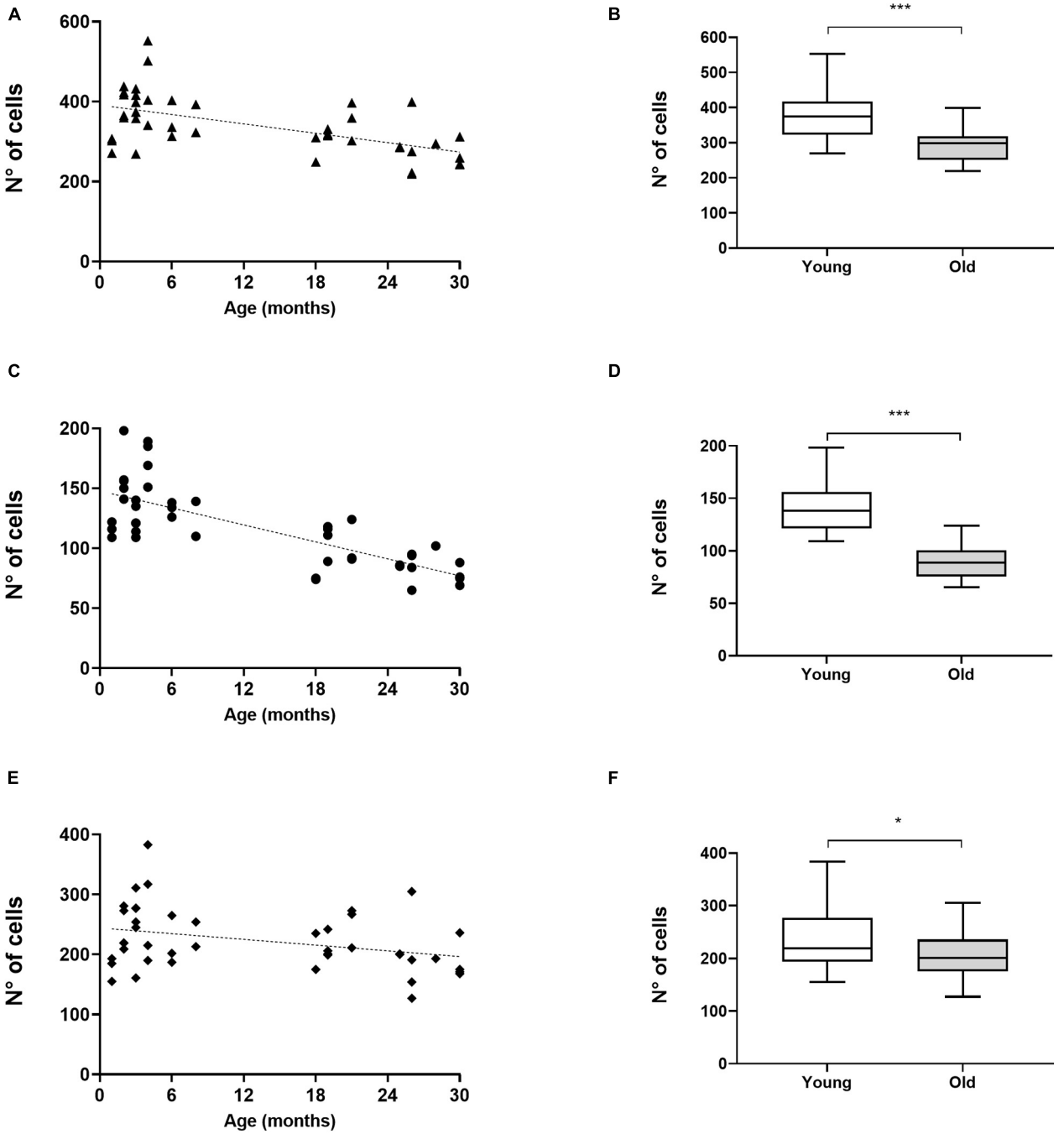

FIGURE 2 | Aging-associated decrease in number of OC cells predominantly affects MOC cells. The image shows the comparisons in the number of CHAT-immunoreactive cells observed in the SOC of young (n = 23) and aged (n = 20) animals. The figures in the left column show the number of cells for each of the animals, where each black symbol represents an individual animal and the dotted line the fitted line. The total number of OC cells is shown in panel **(A)**, while in panel **(C)** is the total number of MOC cells counted in the VNTB and in panel **(E)** the number of LOC cells in the LSO is shown. Linear regressions in panels **(A,C,E)** showed statistical significance (**A:** R-square = 0.3344, F = 20.60, DFn = 1, DFd = 41, $p < 0.0001$; **C:** R-square = 0.5809, F = 56.84, DFn = 1, DFd = 41, $p < 0.0001$; **E:** R-square = 0.1101, F = 5.074, DFn = 1, DFd = 41, $p < 0.0297$). The figures in the right column show boxplots comparing the numbers of CHAT-positive cells between young and aged animals. Panel **(B)** shows the comparison for all OC cells, **(D)** for MOC and **(F)** LOC cells. A significant decrease in cell numbers was found in aged animals for OC cells **(B)** MOC **(D)** and LOC cells **(F)** (Unpaired t test with Welch's correction, **B:** $p < 0.0001$, t = 4.433, df = 39.96; **D:** $p < 0.0001$, t = 7.611, df = 37.99; **F:** p = 0.0409, t = 2.112, df = 40.59). CHAT, choline acetyltransferase; LOC, lateral olivocochlear system; LSO, lateral superior olive; MOC, Medial olivocochlear system; OC, olivocochlear system; SOC, superior olivary complex; VNTB, ventral nucleus of the trapezoid body. The p-values are defined as: $^{*}p < 0.05$ and $^{***}p < 0.0001$.

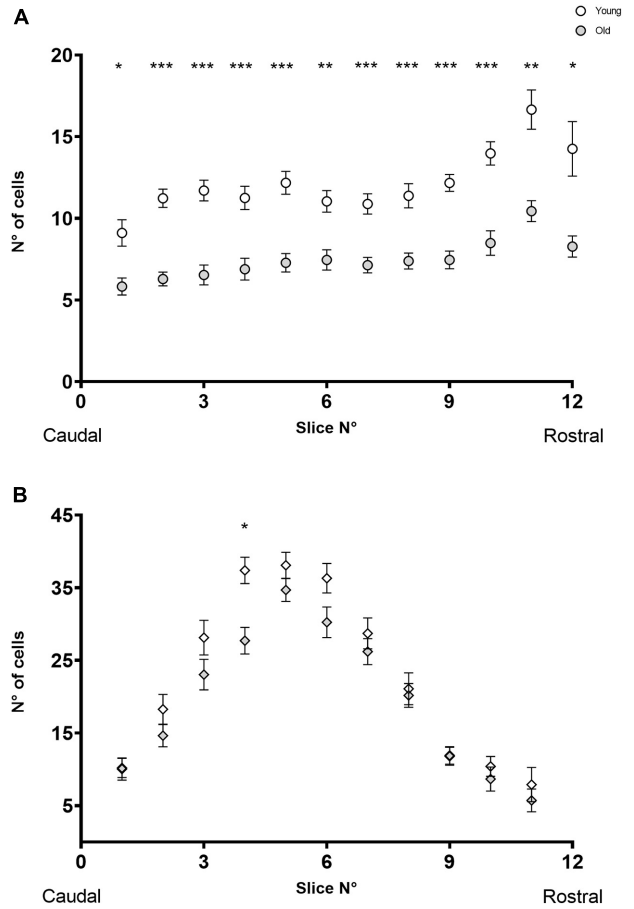

FIGURE 3 | The image shows the number of CHAT-immunoreactive cells found as a function of rostro-caudal localization in the SOC in young (n = 23) and old (n = 20) animals. In panel **(A)** is pictured the distribution of MOC cells, where the empty circles represent the young mice, and the gray circles represent the old animals. The image in panel **(B)** shows the distribution for the LOC cells, where the empty rhombus represents the young mice, and the gray rhombus represent the old ones. A mixed effect analysis found significant differences in the distribution of MOC cells of young vs. old mice (p = < 0.0001, F (1, 41) = 116.0) and the subsequent Sidak multiple comparisons test found significant differences in all the slides. In the case of LOC cells, a mixed effect analysis found significant differences in the distribution of MOC cells of young vs. old mice (p = < 0.0456, F (1, 41) = 4.251) and a Sidak multiple comparisons test found significant differences in slide number 4 (p = 0.0063). Individual points show the mean value, and the error bars indicate the standard error of the mean. Each slide had a width of 50 μm. CHAT, choline acetyltransferase; LOC, lateral olivocochlear system; MOC, medial olivocochlear system; SOC, superior olivary complex. The p-values are defined as: $^{*}p < 0.05$, $^{**}p < 0.001$ and $^{***}p < 0.0001$.

cell number between male and female mice. We found no significant differences between sexes, so we pooled together the male and female data. The number of cells counted in young animals were within the range previously described in mice (Brown and Levine, 2008), 378.3 ± 69.10 for total OC cells, 139.6 ± 26.26 for MOC and 238.1 ± 54.92 for LOC. When comparing young and old animals, significant decreases in the total number of OC cells were observed in the aged animals (**Figures 2A,B**). We found a significant correlation between the age of the animals and the number of OC neurons (**Figure 2A**, R-square = 0.3344, F = 20.60, DFn = 1, DFd = 41, $p < 0.0001$), with aging associated with decreased number of cells. When comparing the young and old mouse groups, we found a 21% significant decrease in the average number of CHAT-positive neurons in the aged mice (**Figure 2B**, $p < 0.0001$, t = 4.433, df = 39.96). When performing these comparisons between young and old animals for the two subdivisions of the OC system, we found that both cell populations showed significant age-associated decreases (**Figures 2C–F**). However, the magnitude of this reduction was not uniform. There was a significant correlation between increasing age and a decrease in number of MOC cells (**Figure 2C**, R-square = 0.5809, F = 56.84, DFn = 1, DFd = 41, $p < 0.0001$), with a 36% average decrease in old mice (**Figure 2D**, $p < 0.0001$, t = 7.611, df = 37.99). The reduction in the number of LOC cells as a function of age was less pronounced (**Figure 2E**, R-square = 0.1101, F = 5.074, DFn = 1, DFd = 41, $p < 0.0297$), with an average reduction of 13% in the number of LOC cells in the old mice (**Figure 2F**, p = 0.0409, t = 2.112, df = 40.59). Thus, these results indicate that the reduction of total number of OC cells during normal aging is mainly driven by a decrease in the number of MOC neurons.

To investigate whether cell losses were restricted to specific regions along the rostro-caudal axis, or if there were group differences that could not be appreciated only by considering the total number of cells per specimen, we compared the distribution of MOC and LOC cells along the rostro-caudal axis in young and old animals. **Figure 3** shows the number of MOC (**Figure 3A**) and LOC (**Figure 3B**) cells as a function of the rostro-caudal axis. For both neuronal populations we found no significant age-associated changes in the pattern of cell distribution in the rostro-caudal axis (Mixed-effects analysis with no significant interaction between age and distribution). In the case of MOC cells, a Sidak's multiple comparisons test between young and old mice yielded significant differences in all observed slides (**Figure 3A**, with p values between 0.04 and <0.0001), indicating that there was a uniform loss of neurons from old animals along the entire axis. In contrast, significant differences were only found in one section of the rostro-caudal LOC axes (**Figure 3B**, slide number 4, Sidak's multiple comparisons test p = 0.0063). These results reinforce the observation that age-related OC neuronal loss is dominated by MOC cells.

Auditory Brainstem Response

In addition to quantifying age-related changes in OC cell number, we also evaluated the auditory responses of young (n = 8, 3 months) and old animals (n = 20, 18–30 months) with ABRs in quiet conditions and in the presence of background noise (BBN of 40 dB SPL, **Figures 4**, **5**). The ABR waveforms (**Figures 4A,B**) and thresholds measured in our animals tested in silence (**Figure 5A**) were consistent with what is expected for CBA/CaJ mice of those ages (Kobrina et al., 2020). Regarding the wave amplitudes, we found significant reduction in the peak to trough amplitude of the ABR waves 1–3 in presence of a masking noise compared to quiet background, both for young (**Figure 4C**; two-way ANOVA

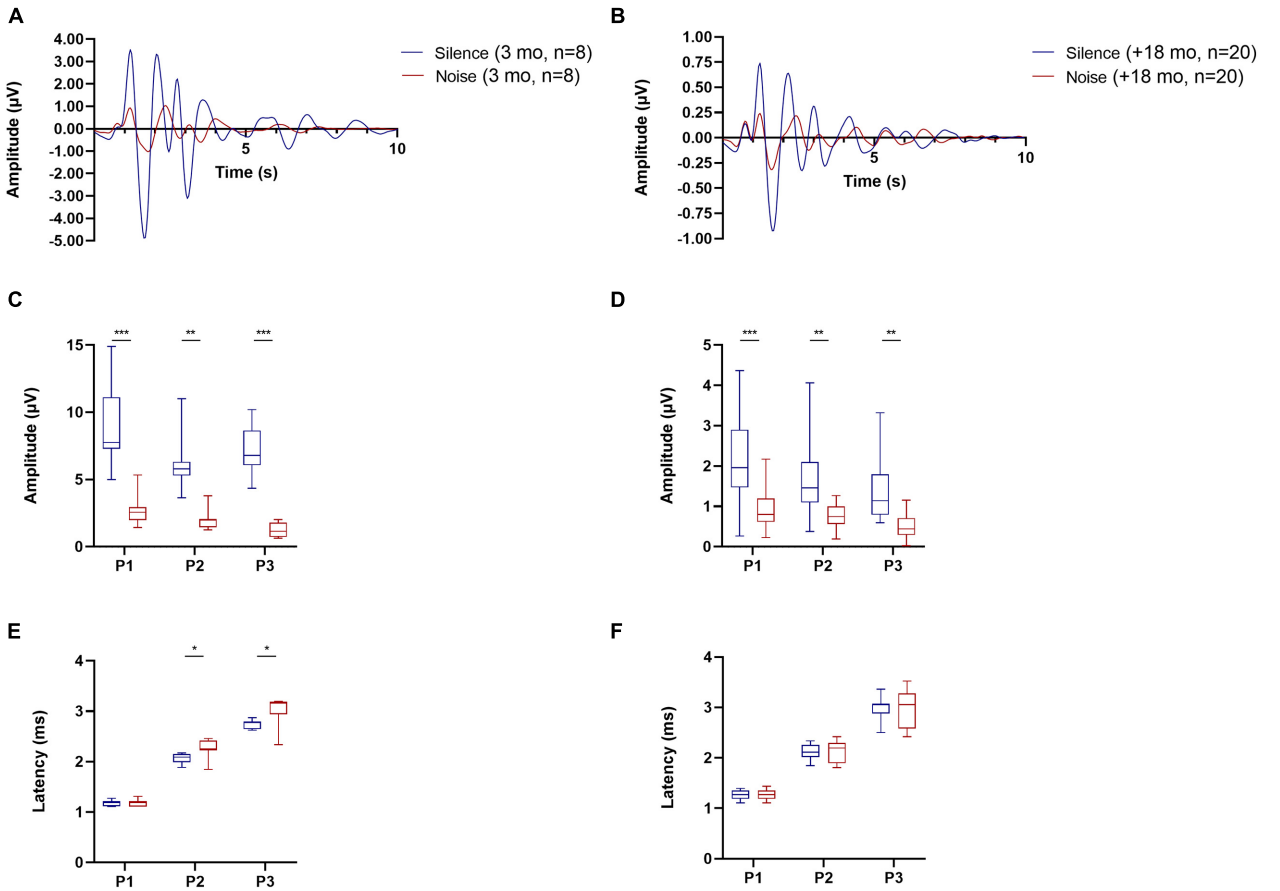

FIGURE 4 | Auditory brainstem response (ABR) waves for young (3 months, n = 8) and old mice between (18–30 months, n = 20) under conditions of silence and background noise. The average ABR waveforms recorded at a 90 dB SPL click for young mice is shown in panel **(A)** and in panel **(B)** for old mice. Blue lines represent the mean values in μV under quiet conditions and the red lines the mean in the presence of 40 dB SPL background noise. Panels **(C,D)** shows box and whiskers plots representing the amplitude for ABR waves 1–3 for young **(C)** and old **(D)** mice. The box and whiskers graph in panels **(E,F)** represent the latency for the peak of ABR waves 1–3 for young and old mice, respectively. Blue boxes show the results in quiet conditions and the red ones under noise (40 dB SPL). Asterisks signal the presence of significant differences between quiet and noise (two-way ANOVA, with a Sidak's multiple comparisons test). The p-values are defined as: $^{*}p < 0.05$, $^{**}p < 0.001$ and $^{***}p < 0.0001$.

$p < 0.0001$, F (1, 42) = 98.91; Sidak's multiple comparisons $p < 0.0001$) and old mice (**Figure 4D** bottom; two-way ANOVA $p < 0.0001$, F (1, 114) = 60.84; Sidak's multiple comparisons $p < 0.0001$). In terms of the ABR wave peak latencies, in young animals we found an increase in the latencies for waves 2 and 3 in the presence of noise compared to quiet (**Figure 4E** bottom; two-way ANOVA $p = 0.0013$, F (1, 42) = 11.94; Sidak's multiple comparisons $p = 0.0443$ for wave 2 and $p = 0.0039$ for wave 3). In the case of the old animals, we found no significant differences in wave 1–3 latencies between the quiet and noise conditions.

Auditory brainstem response thresholds for the young animals ranged between 20.0 ± 5.3 dB SPL for clicks to 27.5 ± 7.1 dB SPL for 8 kHz stimuli (**Figure 5A** pink circles). The threshold values for the old animals were distributed in a range from 36.5 ± 9.3 dB SPL for click stimuli to 60.0 ± 15.4 dB SPL for 24 kHz stimuli (**Figure 5A** gray circles). We found significant differences between young and old animals in thresholds in silence for all stimuli (**Figure 5A**; two-way ANOVA $p < 0.0001$, F (1, 26) = 49.49; Sidak's multiple comparisons test with all values of $p < 0.0001$). When considering ABRs in the noise conditions (**Figure 5B**) we observed that in young mice thresholds ranged between 42.5 ± 7.1 dB SPL for clicks to 58.75 ± 5.2 dB SPL for 12 kHz stimuli (**Figure 5B** pink circles). In addition, when compared to the silent condition, this increase in threshold was significant (two-way ANOVA $p < 0.0001$, F (1, 14) = 125.1; Sidak's multiple comparisons test with all values of $p < 0.0001$). In the case of the old mice, the noise thresholds ranged from 58.0 ± 9.8 dB SPL for click stimuli to 80.5 ± 10.0 dB SPL for 24 kHz stimuli (**Figure 5B** gray circles). This increase in threshold was also statistically significant when compared to the silent condition (two-way ANOVA $p < 0.0001$, F (1, 38) = 67.95; Sidak's multiple comparisons test with all values of $p < 0.0001$). Furthermore, when contrasting the thresholds of young and old animals in the presence of noise, we found that for all stimuli, old animals presented significantly higher thresholds than young animals (**Figure 5B**; two-way ANOVA $p < 0.0001$, F

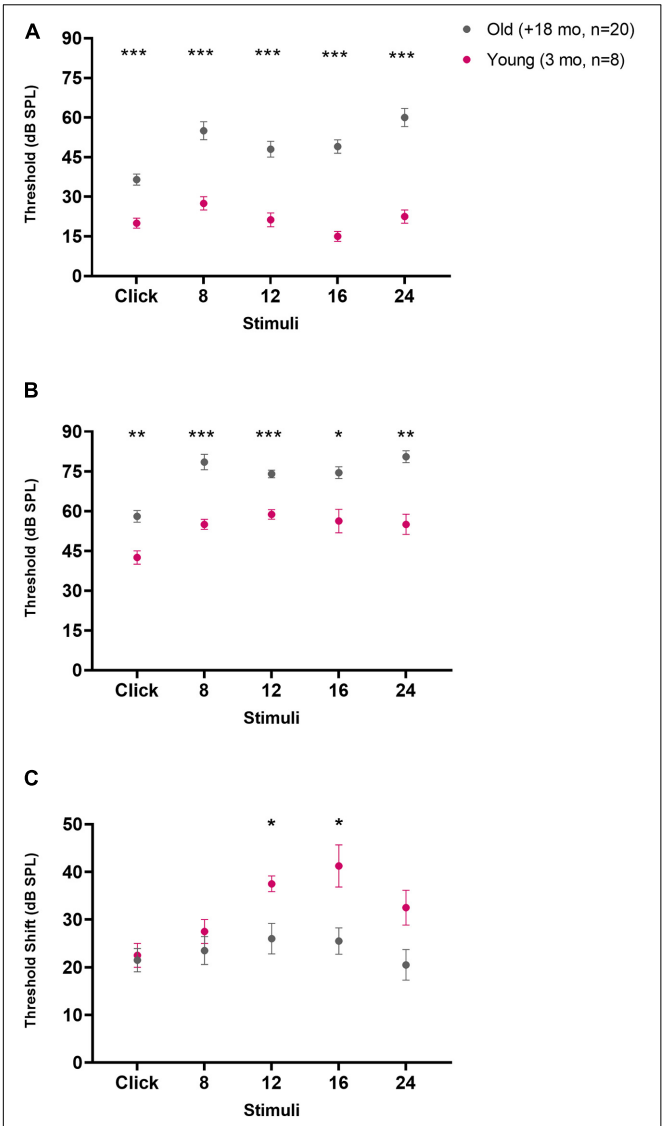

FIGURE 5 | Auditory brainstem response (ABR) thresholds for young (3 months, n = 8) and old mice between (18–30 months, n = 20) under conditions of silence and background noise. Panels **(A,B)** shows the ABR thresholds for young and old mice in quiet and noise conditions, respectively. In panel **(C)** the threshold shift between quiet and noise for young and old mice is shown. The pink circles represent the average threshold value for young animals and the gray circles the average threshold value for the old mice. The error bars show the standard error of the mean. Asterisks signal the presence of significant differences between the young and old animal groups (two-way ANOVA, with a Sidak's multiple comparisons test). The p-values are defined as: $^{*}p < 0.05$, $^{**}p < 0.001$ and $^{***}p < 0.0001$.

(1, 26) = 51.99; Sidak's multiple comparisons with values between p = 0.0181 and p < 0.0001). It is interesting to highlight that the magnitude of the change in thresholds between quiet and noise was not the same for young and old animals (**Figure 5C**). In the presence of noise, the ABR thresholds for young animals had an average increase of 32.25 ± 7.5 dB SPL, whereas in old animals the average increase was 23.40 ± 2.4 dB SPL. We found significant differences between the silence-to-noise threshold shift between the young and old mice group, observing a greater shift in threshold for young animals at the 12 and 16 KHz stimuli (**Figure 5C**; two-way ANOVA p = 0.0291, F (1, 26) = 5.337; Sidak's multiple comparisons test p = 0.0181 for 12 KHz and p = 0.0481 for 16 KHz).

Next, we evaluated whether there were correlations between ABR thresholds in quiet and noise and the number of OC neurons in old mice (**Figure 6**). Considering that the MOC system is involved in hearing in noise, and that one of the main symptoms of ARHL is an impairment of this ability, we were particularly interested in observing whether there was a correlation between ABR thresholds in noise and the number of MOC cells. Notably, we found that animals with higher numbers of MOC neurons tended to present lower click ABR thresholds in noise conditions (**Figure 6A**, R-square = 0.2128, F = 4.896, DFn = 1, DFd = 18, p = 0.0401). In contrast, we found no significant trends with the number of MOC cells for ABRs in silence, nor in any condition analyzed as a function of LOC cell number (**Figures 6C,D**). We also performed these same analyses in young mice (n = 8, 3 months), where we found no correlation between ABR threshold values (in quiet and noise) and the number of olivocochlear cells (MOC and LOC).

In addition to this, we also evaluated the potential associations between ABR wave amplitude (wave 1–3) in quiet and noise conditions with the number of OC neurons. For the old animal group, waves 1 and 2 showed no correlations with MOC cells. But, for wave 3, we found that mice with higher numbers of MOC cells also had higher ABR wave 3 amplitudes in background noise conditions (**Figure 7A** R-square = 0.2577, F = 6.248, DFn = 1, DFd = 18, p = 0.0223) but not in quiet (**Figure 7C**). In contrast, we found no significant correlation for the case of LOC cells (**Figures 7B,D**). For the case of young animals (n = 8, 3 months) we found no correlations in any condition.

Alterations in Synaptic Markers

Complementary to the previous observations, we studied potential age-associated changes in the VNTB and LSO of two synaptic markers: (i) one GABAergic (GAD65) and (ii) one glutamatergic (VGLUT1) (**Figure 8**). We decided to use GAD65 because, under our previous experience, this antibody strongly labels synaptic terminals (Schrode et al., 2018). We quantified the fraction of the area in our ROI that was stained with VGLUT1 or GAD65 in 10 young animals (2–3 months) and 12 old animals (18–28 months). It is important to clarify that this was a separate group from the CHAT-processed animals and, furthermore, that the immunolabeling was not performed on the same mice (five young and six old were used for each condition). These specimens were part of our laboratory archives, obtained from mice used in ABR and cochlear anatomy experiments reported in Kobrina et al. (2020). The first column of **Figure 8** shows the results obtained for VNTB (**Figures 8A,C,E**) while the second column shows the values for LSO (**Figures 8B,D,F**). Within the VNTB we found a significant age-associated decrease in the labeled area for GAD65 (**Figure 8A**, p = 0.0081, t = 3.401, df = 8.850). We did not find significant changes in the labeled area for GAD65 in the LSO (**Figure 8B**), nor for any of the regions with VGLUT1 (**Figures 8C,D**).

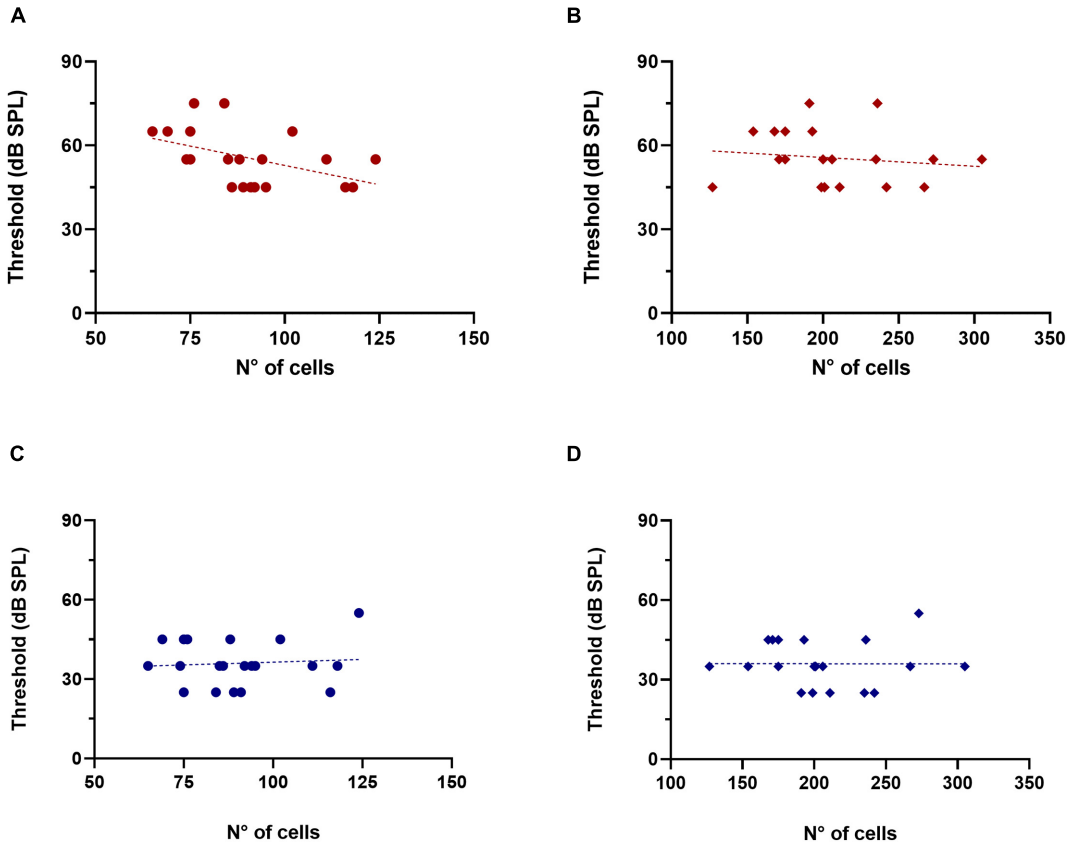

FIGURE 6 | Association between click ABR thresholds in noise and the number of MOC cells in old animals. The figure shows the values of click ABR thresholds in the presence of noise (red) and in silence (blue) as a function of the number of MOC (**A,C**, circles) and LOC cells (**B,D**, rhombuses). The dotted lines correspond to the fitting curve. A significant correlation was found between the number of MOC cells and ABR thresholds in the presence of background noise [**(A)** R-square = 0.2138, $F = 4.896$, DFn = 1, DFd = 18, $p = 0.0401$]. LOC, lateral olivocochlear system; MOC, medial olivocochlear system.

To estimate age-associated changes in the relative amount of excitatory to inhibitory markers in our ROI, we calculated the ratio between the GAD65- and VGLUT1-labeled area for young and adult animals (**Figures 8E,F**). In the case of the VNTB we found a significant decrease in the GAD65/VGLUT1 ratio in old mice (**Figure 8E**, $p = 0.0192$, $t = 2.881$, df = 8.507). In the case of the LSO, we did not find any significant differences. This evidence suggests that during aging, in addition to a loss in neuronal number, the VNTB and the MOC system undergo alterations in the balance of inhibitory and excitatory signals.

Comparison With Other Brainstem Efferent Nuclei

To evaluate whether the observed changes were specific to OC neurons or reflect degeneration of cholinergic neurons in brainstem more generally, we quantified the differences in the number of CHAT-immunoreactive cells in young ($n = 10$) and old ($n = 10$) mice in two other brainstem efferent regions: (i) the trigeminal motor nucleus and (ii) the vestibular efferent nucleus (**Figures 9A,B**, respectively). Although both regions showed slight decreases in the average number of stained cells in old animals (8% for the trigeminal and 7% for the vestibular efferent), we did not find significant differences between the number of cells in young and old animals in either region. This again reinforces the idea that the MOC system could be showing a distinct and significantly greater reduction in cell numbers than other brainstem efferent regions.

DISCUSSION

In this study, we quantified the loss of mouse olivocochlear neurons as a function of age, coupled with ABR measurements in silence and noise. Additionally, we quantified age-associated changes in one GABAergic and one glutamatergic synaptic marker in the LSO and VNTB. We found significant age-associated alterations specific to the MOC system, including (i) a 36% decrease in the number of CHAT-labeled MOC neurons, (ii) a correlation between MOC cell number and ABR threshold in noise, and (iii) a decrease in the GAD65/VGLUT1 ratio in the VNTB. In contrast, the changes we found in LOC were of a smaller extent. We found a lesser decrease in the number of LOC cells (12%) which was not associated with changes in the ABR response. In addition, we also did not find significant changes in the number of CHAT-labeled cells in the vestibular efferent and

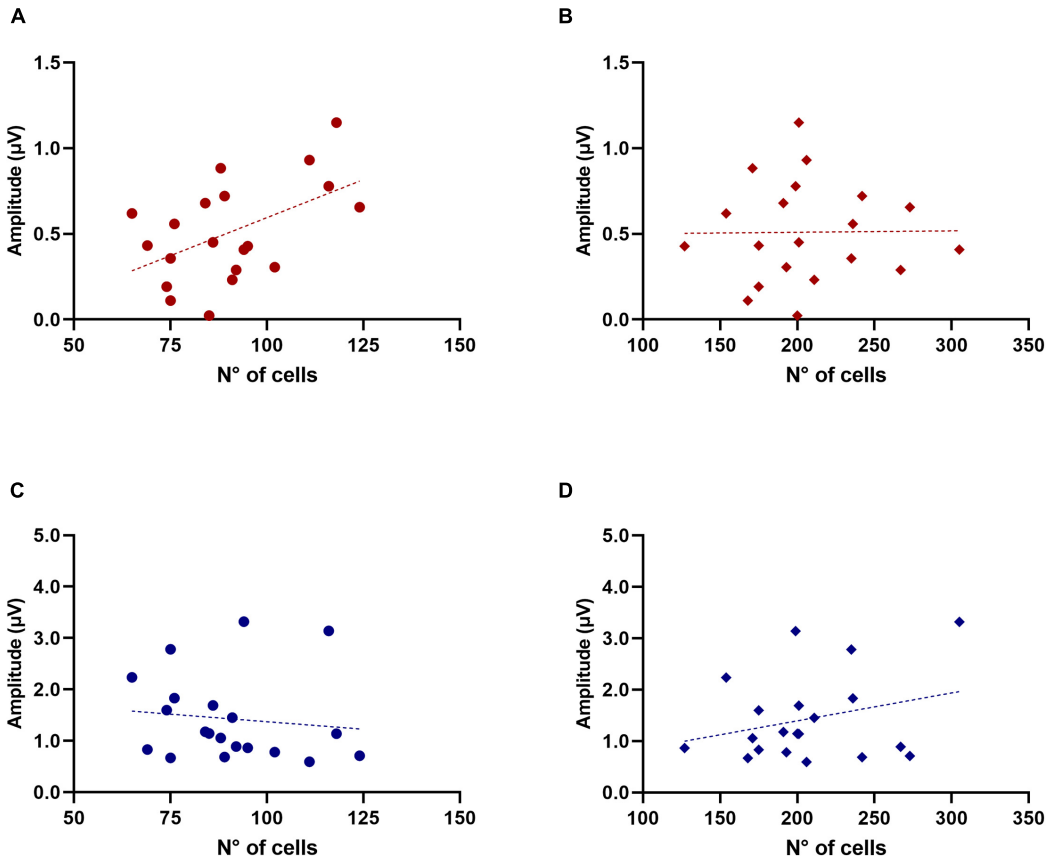

FIGURE 7 | Association between Wave 3 amplitude in noise and the number of MOC cells in old animals. The figure shows ABR wave 3 amplitude in the presence of noise (red) and in silence (blue) as a function of the number of MOC (**A,C**, circles) and LOC cells (**B,D**, rhombuses). The dotted lines correspond to the fitting curve. A significant correlation was found between the number of MOC cells and wave 3 amplitude in the presence of background noise [**(A)** R-square = 0.2577, $F = 6.248$, DFn = 1, DFd = 18, $p = 0.0223$]. LOC, lateral olivocochlear system; MOC, medial olivocochlear system.

trigeminal motor nucleus neurons. These observations indicate differential vulnerability to aging of MOC cells and provide evidence for their role in ARHL.

The results we obtained regarding the number of OC, trigeminal, and vestibular cells in young animals were within the expected ranges for mice (Sturrock, 1987; Campbell and Henson, 1988; Brown and Levine, 2008; Mathews et al., 2015). In old animals, the loss of OC neurons was dominated by MOC loss (**Figures 2D**, **3A**) being on average 2.7 times greater than the loss of LOC cells. This decline in neuronal number was symmetrical between both hemispheres and along the rostro-caudal axis, with little inter-subject variability and no significant differences between male and female mice. Such differential reduction led to changes in the ratio of LOC to MOC cells in the OC system, from an average of 1.7:1 LOC per MOC in young mice to 2.3:1 in old mice.

These findings are congruent with previous evidence observed in studies that have explored age-related OC alterations in mice and humans. Age-associated decreases in contralateral suppression of distortion product otoacoustic emission levels have been observed in humans and CBA mice (Jacobson et al., 2003). Considering that this response is mediated by the MOC system, even if middle ear muscles contribute to the effect (Xu et al., 2017; Valero et al., 2018), this is consistent with age-related MOC dysfunction. Moreover, anatomical research in mice and humans also supports these differential changes. There is a loss in the density of cochlear MOC neurons as a function of age, while the density of LOC innervation does not decrease overall (Liberman and Liberman, 2019; Jeng et al., 2020; Kobrina et al., 2020). However, other age-associated changes in LOC synapses have been identified. There is a synaptic rearrangement involving an increase in efferent innervation of IHCs in mice with accelerated ARHL, in which the aging cochlea recovers some features of postnatal development with the re-emergence of efferent inhibition of the IHCs (Lauer et al., 2012; Zachary and Fuchs, 2015). All this evidence points to a loss of MOC neurons and synapses with the OHC, accompanied by changes in LOC synaptic organization without substantial loss of neuronal cell bodies.

Moreover, the absence of observed changes in the number of CHAT-labeled cells in the trigeminal motor nucleus, vestibular efferents, and LOC also is in line with the changes expected in the aging brain, where we would not anticipate seeing large reductions overall in the number of neurons, except in

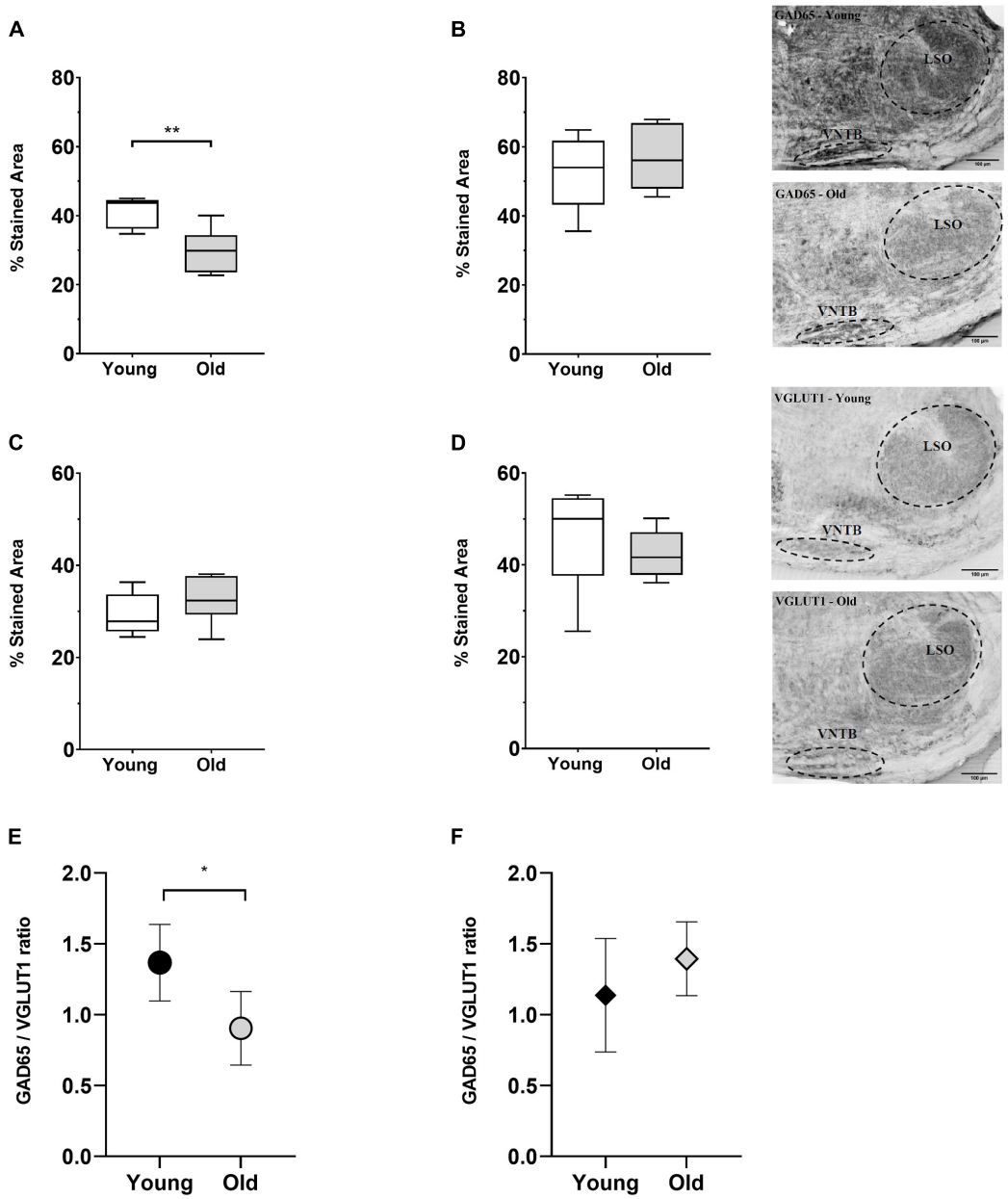

FIGURE 8 | Age-associated changes of inhibitory and excitatory synaptic markers. The first row shows boxplots representing the fraction of GAD65-immunoreactive area of young (n = 5) and old mice (n = 6, gray-filled boxes) in the VNTB **(A)** and LSO **(B)**. For the VNTB in **(A)**, a significant difference was found between young and old mice (unpaired t test with Welch's correction, p = 0.0081, t = 3.401, df = 8.850). The images on the right correspond to an example of GAD65 staining at 10x in a young male mouse (top, 3 m.o.) and an old male mouse (bottom, 26 m.o.). The second row is showing boxplots that represent the fraction of VGLUT1-immunoreactive area of young (n = 5) and old mice (n = 6, gray-filled boxes) in the VNTB **(C)** and LSO **(D)**. On the right, there are two examples of staining of VGLUT1 at 10x for a young female mouse (top, 2 m.o.) and an old female mouse (bottom, 28 m.o.). The last row shows a comparison of the GAD65/VGLUT1 ratio between young and old animals, where **(E)** represents the VNTB and **(F)** the LSO, and the error bars indicate the standard error of the mean. In the case of VNTB **(E)**, a significant difference was found between the two groups (unpaired t test with Welch's correction, p = 0.0192, t = 2.881, df = 8.507). GAD65, glutamic acid decarboxylase 65-kilodalton isoform; LSO, lateral superior olive; VGLUT1, vesicular glutamate transporter 1; VNTB, ventral nucleus of the trapezoid body. The p-values are defined as: $^{*}p < 0.05$ and $^{**}p < 0.001$.

specific regions (Morrison and Hof, 1997; Pannese, 2011). There are interesting similarities between our results and findings from research conducted in aging gerbils that also quantified the number of OC, trigeminal and vestibular efferent cell bodies (Radtke-Schuller et al., 2015). This study also found that the decrease in cell number was restricted to the OC system, with no alterations in other efferent pathways. This suggests a greater vulnerability to aging of the OC system

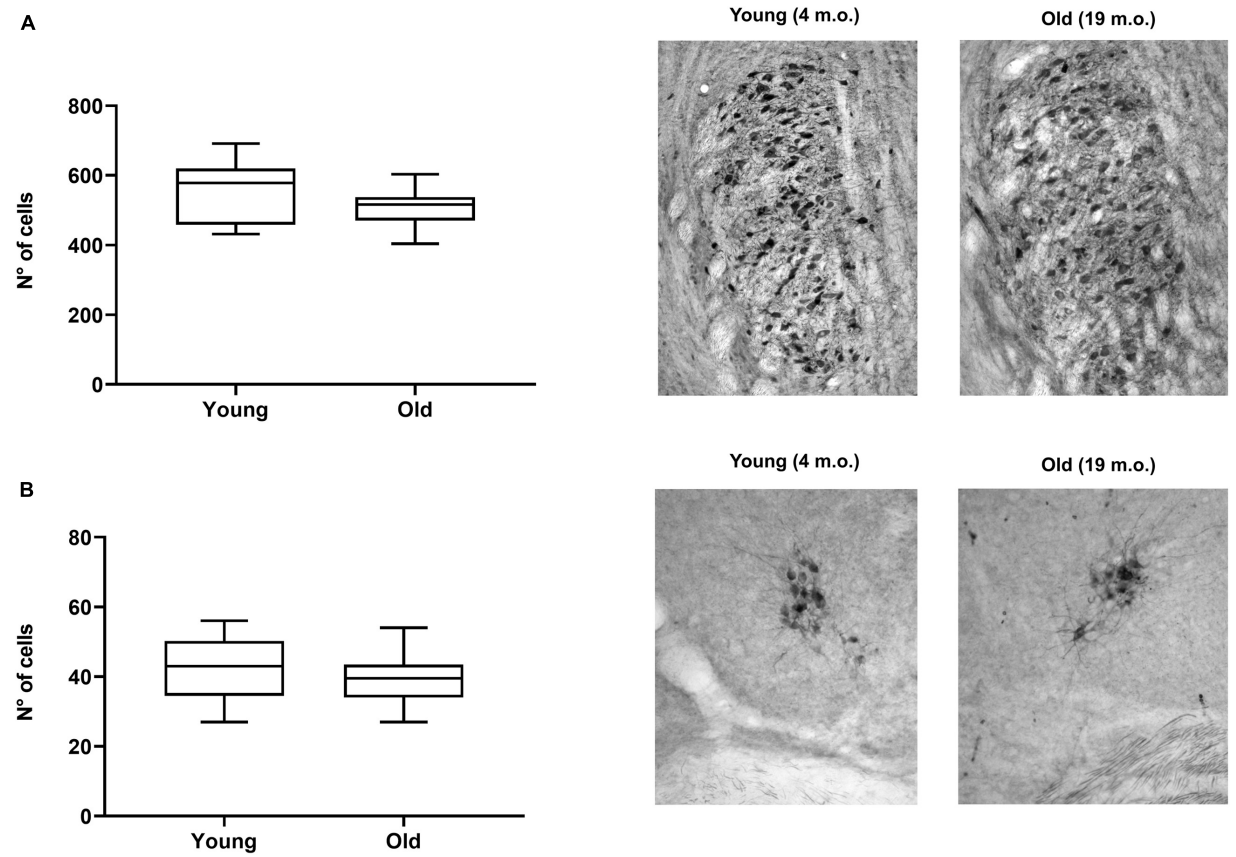

FIGURE 9 | Comparison of trigeminal and vestibular efferent cell numbers in young (n = 10) and aged mice (n = 10). The figure shows boxplots comparing the number of CHAT-positive cells between young and aged animals for the trigeminal motor nucleus **(A)** and the vestibular efferent nucleus **(B)**. The images on the right of each boxplot correspond to photomicrographs examples of trigeminal **(A)** and vestibular efferent **(B)** sections stained with acetylcholinesterase for young and old mice. We found no significant differences between young and old mice in neither of the two cases. CHAT, choline acetyltransferase.

compared to other efferent regions of the brainstem. However, for gerbils the OC decrease was distributed in both MOC and LOC subpopulations, without changes in the ratio of MOC and LOC cells in old animals. This species difference in the vulnerability to OC neurons could be related to the particular susceptibility of gerbils to auditory brainstem degeneration from exposure to environmental noise over the course of a lifetime (Statler et al., 1990; McGinn and Faddis, 1994, 1998; Gleich and Strutz, 2002).

The mechanisms underlying the vulnerability of the OC system and, especially the MOC, cannot be directly concluded from our results, but there are at least three factors that could be contributing: (i) differential alterations in the brainstem regions, (ii) a specific vulnerability of MOC cells, or (iii) differential alterations in peripheral synapses.

For the first factor, our data describing the age-dependent changes in the fraction of GAD65 and VGLUT1 suggest that VNTB shows alterations that are not observed in LSO (**Figure 8**). The decrease in the fraction of GAD65 and consequently the GAD65/VGLUT1 ratio could be reflecting a decline in the proportion of inhibitory inputs in the region. These results are also in line with previous publications reporting an age-related decrease in GABA and GAD levels in other central regions of the auditory system (Caspary et al., 1990; Burianova et al., 2009; Ibrahim and Llano, 2019). In addition to the functional effects that may result from the decrease in GABA-mediated inhibition (Caspary et al., 2008), there is also the possibility that it may be contributing to the vulnerability of VNTB, potentially through excitotoxicity effects.

On the other hand, the absence of observed changes in the LSO is consistent with previous results in gerbils, where no age-related effects on GABAergic and glycinergic markers were found (Gleich et al., 2004). However, our data has several caveats that limit our interpretation. The GAD65 and VGLUT1 measurements did not come from the same animals, although the animals were raised and aged in the same environment during the same time periods, and the specimens were prepared by the same experimenter. Also, we did not capture all the excitatory and inhibitory inputs, for instance those that are positive for VGLUT2 instead of VGLUT1. It is also unclear if the GAD65 only represents GABAergic or also GAD-positive glycinergic inputs. Future experiments will tease apart the contributions of different sources of inputs to OC neurons to the observed age-related effects.

Regarding the specific vulnerability of MOC neurons, changes in somatic expression levels of the Kv3.1 channel have been observed in MOC cells of aged mice (Zettel et al., 2007), suggesting a higher vulnerability than their LOC counterparts. Complementary to this, extracellular recordings from the LSO in rats have shown that neuronal populations do not significantly change their responses with age, supporting the idea of their resistance to degeneration (Finlayson and Caspary, 1993). Another notable difference between MOC, LOC and vestibular efferent cells is the presence of myelin sheath in MOCs. While intuition suggests that myelin should be a protective factor in the face of trauma and aging (Ceballos et al., 1999; Reeves et al., 2005), there are situations where it has been observed to be more vulnerable (Fujimura et al., 1991; Tang et al., 1997). It is also possible that the loss of MOC neurons is related to other factors such as a higher energy demand, which the sole presence of myelin is not necessarily a good predictor (Harris and Attwell, 2012).

There is evidence supporting the idea that changes at the level of the cochlea may be contributing to the observed differences in loss of LOC and MOC neuron cell bodies. For example, the modifications in the number of MOC efferent neurons could be explained by degeneration and decrease in the number of OHCs, whereas IHCs remain much more intact with age (Jeng et al., 2020; Kobrina et al., 2020; Wu et al., 2020). In addition, aged OHCs manifest morphological alterations, such as a reduction in size, that precede the decrease in the number of MOC efferent fibers (Jeng et al., 2020). In contrast, IHC numbers do not decrease with age to the same extent as OHCs do with age (Spongr et al., 1997; Sergeyenko et al., 2013; Kobrina et al., 2020). Rather, there is a subpopulation of IHCs that reverts into an immature-like biophysical and morphological profile (Corns et al., 2018), which are possibly be the ones that are re-innervated by inhibitory cholinergic efferents (Lauer et al., 2012; Corns et al., 2018).

Regarding our ABR results, the thresholds we identified in young and old animals were within the expected ranges, as were the wave 1–3 latencies and amplitudes (Henry, 1979; Schrode et al., 2018; Kobrina et al., 2020). In this context, the observed decrease in ABR amplitude in old mice compared to young mice indicates the existence of a reduced physiological response to sound. The absolute increase in the ABR threshold in quiet condition is also a reflection of this (**Figure 5**). When looking at our results from shifts in thresholds between the quiet and noise it is noticeable that older animals had smaller shifts than the young adults. This result is consistent with previous findings in mice and may be reflecting OHC damage or ceiling effects due to the already increased thresholds (Ehret, 1979). However, inferring the exact cause of these findings would be extremely challenging. Previous evidence shows that the changes associated with ARHL in these mice are multifactorial, which makes it difficult to attribute a single main cause of ARHL to a particular type of cell degeneration. For example, Kobrina et al. (2020) showed that multiple alterations at the cochlear level, including loss of hair cells, synaptic ribbons, and changes in the stria vascularis are related to ARHL.

One of the most remarkable findings was the correlation between the number of MOC cells in old animals with the ABR thresholds (**Figure 6A**) and with the wave 3 amplitudes (**Figure 7A**) in noise conditions. These correlations were not found in young animals. There is previous evidence showing that the MOC system can suppress the cochlear response to continuous noise and, thereby, it can help unmask transient tone stimuli, making them more detectable (Winslow and Sachs, 1988; Liberman and Guinan, 1998). Therefore, it is reasonable to have found that a higher number of MOC cells correlated with lower ABR thresholds in noise. It is also worthy of note that the superior olive is directly related with wave 3 (Buchwald and Huang, 1975; Henry, 1979; Melcher et al., 1996) and, that P3 and N3 are lost with the application of local anesthetic into the trapezoid body (Wada and Starr, 1983). Given that this manipulation results in non-specific effects in the trapezoidal body, the precise contribution of MOC neurons is unknown. However, it does show that diminished function in the vicinity of the MOC cell bodies reduces this ABR component, which supports the idea that this region is altered in aging. On the other hand, it is also reasonable that no correlations were observed with waves 1 and 2 of the ABR. The amplitude of wave 1 is dominated by age-related cochlear degeneration that is well documented in this strain, including in our laboratory (Kobrina et al., 2020). In addition, age-related increases in olivocochlear innervation of inner hair cells may further reduce the auditory nerve activity that generates ABR wave 1 (Lauer et al., 2012; Zachary and Fuchs, 2015; Kobrina et al., 2020). Therefore, it is difficult to infer how the loss of MOC inputs would independently contribute to the already diminished wave 1. Centrally generated ABR waves often show non-linear changes in response to peripheral damage and diminished ABR wave 1 (for instance, as shown by our lab in Schrode et al., 2018).

The absence of this correlation with wave 3 in young animals was also reasonably expected. This is due to the fact that the ABR thresholds and the number of neurons were distributed across narrower ranges in young animals. In an aged and impaired system, it is more probable that the decrease in the number of MOC cells has an observable impact on signal processing in noise since redundant mechanisms for hearing in noise may be diminished by age-related cochlear and central degeneration. Other functions such as protection against acoustic trauma and selective attention could also be affected by the decrease in MOC numbers and may contribute to further deterioration of auditory pathway and cognitive functions. Further experiments are needed in order to explore these questions.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The animal study was reviewed and approved by Johns Hopkins University Institutional Animal Care and Use Committee.

AUTHOR CONTRIBUTIONS

AL: original idea, project administration and supervision, and funding acquisition. AL and SV-J: experimental conceptualization and data analysis. SV-J: experimental development and manuscript writing. SV-J and MW: tissue processing. SV-J and GB-M: image processing. AL, SV-J, and GB-M: manuscript editing. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by NIDCD grant R01 DC017620.

ACKNOWLEDGMENTS

We thank Miguel Concha-Miranda and Fernando C Ortiz for their valuable feedback and suggestions on the manuscript and Ryan Rigel for his technical assistance.

REFERENCES

Abercrombie (1946). Estimation of nuclear population from microtome sections. *Anat. Rec.* 94, 239–247. doi: 10.1002/ar.1090940210

Agrawal, Y., Platz, E. A., and Niparko, J. K. (2008). Prevalence of hearing loss and differences by demographic characteristics among us adults: data from the National Health and Nutrition Examination Survey, 1999-2004. *Arch. Intern. Med.* 168, 1522–1530. doi: 10.1001/archinte.168.14.1522

Blanchet, C., Erostegui, C., Sugasawa, M., and Dulon, D. (1996). Acetylcholine-induced potassium current of guinea pig outer hair cells: its dependence on a calcium influx through nicotinic-like receptors. *J. Neurosci.* 16, 2574–2584. doi: 10.1523/JNEUROSCI.16-08-02574.1996

Bobbin, R. P., and Konishi, T. (1971). Acetylcholine mimics crossed olivocochlear bundle stimulation. *Nat. New Biol.* 231, 222–223. doi: 10.1038/newbio231222a0

Boero, L. E., Castagna, V. C., Terreros, G., Moglie, M. J., Silva, S., Maass, J. C., et al. (2020). Preventing presbycusis in mice with enhanced medial olivocochlear feedback. *Proc. Natl. Acad. Sci.U.S.A.* 117, 11811–11819. doi: 10.1073/pnas.2000760117

Bowl, M. R., and Dawson, S. J. (2019). Age-related hearing loss. *Cold Spring Harb. Perspect. Med.* 9:a033217. doi: 10.1101/cshperspect.a033217

Brown, M. C. (2011). "Anatomy of olivocochlear neurons," in *Auditory and Vestibular Efferents*, eds D. K. Ryugo and R. R. Fay (New York, NY: Springer New York), 17–37. doi: 10.1007/978-1-4419-7070-1_2

Brown, M. C., and Levine, J. L. (2008). Dendrites of medial olivocochlear neurons in mouse. *Neuroscience* 154, 147–159. doi: 10.1016/j.neuroscience.2007.12.045

Buchwald, J. S., and Huang, C. (1975). Far-field acoustic response: origins in the cat. *Science* 189, 382L–384. doi: 10.1126/science.1145206

Burianova, J., Ouda, L., Profant, O., and Syka, J. (2009). Age-related changes in GAD levels in the central auditory system of the rat. *Exp. Gerontol.* 44, 161–169. doi: 10.1016/j.exger.2008.09.012

Campbell, J. P., and Henson, M. M. (1988). Olivocochlear neurons in the brainstem of the mouse. *Hear. Res.* 35, 271–274. doi: 10.1016/0378-5955(88)90124-4

Caspary, D. M., Ling, L., Turner, J. G., and Hughes, L. F. (2008). Inhibitory neurotransmission, plasticity and aging in the mammalian central auditory system. *J. Exp. Biol.* 211, 1781–1791. doi: 10.1242/jeb.013581

Caspary, D. M., Raza, A., Lawhorn Armour, B. A., Pippin, J., and Arneric, S. P. (1990). Immunocytochemical and neurochemical evidence for age-related loss of GABA in the inferior colliculus: implications for neural presbycusis. *J. Neurosci.* 10, 2363–2372. doi: 10.1523/JNEUROSCI.10-07-02363.1990

Ceballos, D., Cuadras, J., Verdú, E., and Navarro, X. (1999). Morphometric and ultrastructural changes with ageing in mouse peripheral nerve. *J. Anat.* 195, 563–576. doi: 10.1046/j.1469-7580.1999.19540563.x

Ciuman, R. R. (2010). The efferent system or olivocochlear function bundle - fine regulator and protector of hearing perception. *Int. J. Biomed. Sci.* 6, 276–288.

Collins, J. G. (1997). Prevalence of selected chronic conditions: United States, 1990-1992. *Vital Health Stat.* 10, 1–89.

Cooper, N. P., and Guinan, J. J. Jr. (2006). Efferent-mediated control of basilar membrane motion. *J. Physiol.* 576, 49–54. doi: 10.1113/jphysiol.2006.114991

Corns, L. F., Johnson, S. L., Roberts, T., Ranatunga, K. M., Hendry, A., Ceriani, F., et al. (2018). Mechanotransduction is required for establishing and maintaining mature inner hair cells and regulating efferent innervation. *Nat. Commun.* 9:4015. doi: 10.1038/s41467-018-06307-w

Dubno, J. R., Dirks, D. D., and Morgan, D. E. (1984). Effects of age and mild hearing loss on speech recognition in noise. *J. Acoust. Soc. Am.* 76, 87–96. doi: 10.1121/1.391011

Ehret, G. (1979). Correlations between cochlear hair cell loss and shifts of masked and absolute behavioral auditory thresholds in the house mouse. *Acta Otolaryngol.* 87, 28–38. doi: 10.3109/00016487909126384

Eybalin, M. (1993). Neurotransmitters and neuromodulators of the mammalian cochlea. *Physiol. Rev.* 73, 309–373. doi: 10.1152/physrev.1993.73.2.309

Felix, D., and Ehrenberger, K. (1992). The efferent modulation of mammalian inner hair cell afferents. *Hear. Res.* 64, 1–5. doi: 10.1016/0378-5955(92)90163-H

Finlayson, P. G., and Caspary, D. M. (1993). Response properties in young and old Fischer-344 rat lateral superior olive neurons: a quantitative approach. *Neurobiol. Aging* 14, 127–139. doi: 10.1016/0197-4580(93)90088-S

Flurkey, K., Currer, J. M., and Harrison, D. E. (2007). "The mouse in Aging research," in *The Mouse in Biomedical Research*, eds J. G. Fox, S. W. Barthold, M. T. Davisson, C. E. Newcomer, F. W. Quimby, and A. L. Smith (Burlington, MA: Elsevier Inc), 637–672.

Frisina, D. R., and Frisina, R. D. (1997). Speech recognition in noise and presbycusis: relations to possible neural mechanisms. *Hear. Res.* 106, 95–104. doi: 10.1016/s0378-5955(97)00006-3

Fuchs, P. A., and Lauer, A. M. (2018). Efferent Inhibition of the Cochlea. *Cold Spring Harb. Perspect. Med.* 9:a033530. doi: 10.1101/cshperspect.a033530

Fujimura, H., Lacroix, C., and Said, G. (1991). Vulnerability of nerve fibres to ischaemia: a quantitative light and electron microscope study. *Brain* 114, 1929–1942. doi: 10.1093/brain/114.4.1929

Gleich, O., and Strutz, J. (2002). Age dependent changes in the medial nucleus of the trapezoid body in gerbils. *Hear. Res.* 164, 166–178. doi: 10.1016/S0378-5955(01)00430-0

Gleich, O., Weiss, M., and Strutz, J. (2004). Age-dependent changes in the lateral superior olive of the gerbil (Meriones unguiculatus). *Hear. Res.* 194, 47–59. doi: 10.1016/j.heares.2004.03.016

Gordon-Salant, S. (2005). Hearing loss and aging: new research findings and clinical implications. *J. Rehabil. Res. Dev.* 42, 9–24. doi: 10.1682/jrrd.2005.01.0006

Groff, J. A., and Liberman, M. C. (2003). Modulation of cochlear afferent response by the lateral olivocochlear system: activation via electrical stimulation of the inferior Colliculus. *J. Neurophysiol.* 90, 3178–3200. doi: 10.1152/jn.00537.2003

Guinan et al., J. J., Warr, W. B., and Norris, B. E. (1983). Differential olivocochlear projections from lateral versus medial zones of the superior olivary complex. *J. Comp. Neurol.* 221, 358–370. doi: 10.1002/cne.902210310

Guinan et al., J. J., Warr, W. B., and Norris, B. E. (1984). Topographic organization of the olivocochlear projections from the lateral and medial zones of the superior olivary complex. *J. Comp. Neurol.* 226, 21–27. doi: 10.1002/cne.902260103

Guinan, J. J. (1996). "Physiology of olivocochlear efferents," in *The Cochlea*, eds P. Dallos, A. N. Popper, and R. R. Fay (New York, NY: Springer New York), 435–502. doi: 10.1007/978-1-4612-0757-3_8

Harris, J. J., and Attwell, D. (2012). The Energetics of CNS White Matter. *J. Neurosci.* 32, 356–371. doi: 10.1523/JNEUROSCI.3430-11.2012

Henry, K. R. (1979). Auditory brainstem volume-conducted responses: origins in the laboratory mouse. *J. Am. Aud. Soc.* 4, 173–178.

Ibrahim, B. A., and Llano, D. A. (2019). Aging and central auditory disinhibition: is it a reflection of homeostatic downregulation or metabolic vulnerability? *Brain Sci.* 9:351. doi: 10.3390/brainsci9120351

Jacobson, M., Kim, S., Romney, J., Zhu, X., and Frisina, R. D. (2003). Contralateral suppression of distortion-product otoacoustic emissions declines with age: a comparison of findings in CBA mice with human listeners. *Laryngoscope* 113, 1707–1713. doi: 10.1097/00005537-200310000-00009

Jeng, J.-Y., Johnson, S. L., Carlton, A. J., De Tomasi, L., Goodyear, R. J., De Faveri, F., et al. (2020). Age-related changes in the biophysical and morphological characteristics of mouse cochlear outer hair cells. *J. Physiol.* 598, 3891–3910. doi: 10.1113/JP279795

Kawase, T., and Liberman, M. C. (1993). Antimasking effects of the olivocochlear reflex. I. Enhancement of compound action potentials to masked tones. *J. Neurophysiol.* 70, 2519–2532. doi: 10.1152/jn.1993.70.6.2519

Kim, S. H., Frisina, D. R., and Frisina, R. D. (2002). Effects of age on contralateral suppression of distortion product otoacoustic emissions in human listeners with normal hearing. *Audiol. Neurotol.* 7, 348–357. doi: 10.1159/000066159

Kobrina, A., and Dent, M. L. (2020). The effects of age and sex on the detection of pure tones by adult CBA/CaJ mice (*Mus musculus*). *J. Neurosci. Res.* 98, 1731–1744. doi: 10.1002/jnr.24496

Kobrina, A., Schrode, K. M., Screven, L. A., Javaid, H., Weinberg, M. M., Brown, G., et al. (2020). Linking anatomical and physiological markers of auditory system degeneration with behavioral hearing assessments in a mouse (*Mus musculus*) model of age-related hearing loss. *Neurobiol. Aging* 96, 87–103. doi: 10.1016/j.neurobiolaging.2020.08.012

Lauer, A. M. (2017). Minimal effects of age and exposure to a noisy environment on hearing in alpha9 nicotinic receptor knockout mice. *Front. Neurosci.* 11:1–10. doi: 10.3389/fnins.2017.00304

Lauer, A. M., Fuchs, P. A., Ryugo, D. K., and Francis, H. W. (2012). Efferent synapses return to inner hair cells in the aging cochlea. *Neurobiol. Aging* 33, 2892–2902. doi: 10.1016/j.neurobiolaging.2012.02.007

Lauer, A. M., Jimenez, S. V., and Delano, P. H. (2021). Olivocochlear efferent effects on perception and behavior. *Hear. Res.* 108207. doi: 10.1016/j.heares.2021.108207 [Epub ahead of print].

Le Prell, C. G., Shore, S. E., Hughes, L. F., and Bledsoe, S. C. (2003). Disruption of Lateral Efferent Pathways: Functional Changes in Auditory Evoked Responses. *J. Assoc. Res. Otolaryngol.* 4, 276–290. doi: 10.1007/s10162-002-3018-6

Liberman, L. D., and Liberman, M. C. (2019). Cochlear Efferent Innervation Is Sparse in Humans and Decreases with Age. *J. Neurosci* 39, 9560L–9569. doi: 10.1523/JNEUROSCI.3004-18.2019

Liberman, M. C., Dodds, L. W., and Pierce, S. (1990). Afferent and efferent innervation of the cat cochlea: quantitative analysis with light and electron microscopy. *J. Comp. Neurol.* 301, 443–460. doi: 10.1002/cne.903010309

Liberman, M. C., and Guinan, J. J. (1998). "Feedback control of the auditory periphery: anti-masking effects of middle ear muscles vs. olivocochlear efferents. *J. Commun. Disord.* 31, 471–482 quiz 483; 553. doi: 10.1016/s0021-9924(98)00019-7 quiz 483; 553.

Liberman, M. C., Liberman, L. D., and Maison, S. F. (2014). Efferent feedback slows cochlear Aging. *J. Neurosci.* 34, 4599L–4607. doi: 10.1523/JNEUROSCI.4923-13.2014

Lin, F. R. (2011). Hearing loss and cognition among older adults in the United States. *J. Gerontol. A Biol. Sci. Med. Sci.* 66, 1131–1136. doi: 10.1093/gerona/glr115

Lina, I. A., and Lauer, A. M. (2013). Rapid measurement of auditory filter shape in mice using the auditory brainstem response and notched noise. *Hear. Res.* 298, 73–79. doi: 10.1016/j.heares.2013.01.002

Livingston, G., Sommerlad, A., Orgeta, V., Costafreda, S. G., Huntley, J., Ames, D., et al. (2017). Dementia prevention, intervention, and care. *Lancet* 390, 2673–2734. doi: 10.1016/S0140-6736(17)31363-6

Loughrey, D. G., Kelly, M. E., Kelley, G. A., Brennan, S., and Lawlor, B. A. (2018). Association of age-related hearing loss with cognitive function, cognitive impairment, and dementia: a systematic review and meta-analysis. *JAMA Otolaryngol. Head Neck Surg.* 144, 115–126. doi: 10.1001/jamaoto.2017.2513

Maison, S. F., Adams, J. C., and Liberman, M. C. (2003). Olivocochlear innervation in the mouse: Immunocytochemical maps, crossed versus uncrossed contributions, and transmitter colocalization. *J. Comp. Neurol.* 455, 406–416. doi: 10.1002/cne.10490

Malmierca, M., and Ryugo, D. (2011). "Descending connections of auditory cortex to the midbrain and brain stem," in *The Auditory Cortex*, eds J. Winer and C. Schreiner (New York, NY: Springer), 189–208.

Mathews, M. A., Murray, A., Wijesinghe, R., Cullen, K., Tung, V. W. K., and Camp, A. J. (2015). Efferent vestibular neurons show homogenous discharge output but heterogeneous synaptic input profile in vitro. *PLoS One* 10:e0139548. doi: 10.1371/journal.pone.0139548

May, B. J., Budelis, J., and Niparko, J. K. (2004). Behavioral Studies of the olivocochlear efferent system: learning to listen in noise. *Arch. Otolaryngol. Neck Surg.* 130, 660–664. doi: 10.1001/archotol.130.5.660

McGinn, M. D., and Faddis, B. T. (1994). Exposure to low frequency noise during rearing induces spongiform lesions in gerbil cochlear nucleus: high frequency exposure does not. *Hear. Res.* 81, 57–65. doi: 10.1016/0378-5955(94)90153-8

McGinn, M. D., and Faddis, B. T. (1998). Neuronal degeneration in the gerbil brainstem is associated with spongiform lesions. *Microsci. Res. Tech.* 41, 187–204. doi: 10.1002/(SICI)1097-0029(19980501)41:3<187::AID-JEMT3<3.0.CO;2-R

McGuire, B., Fiorillo, B., Ryugo, D. K., and Lauer, A. M. (2015). Auditory nerve synapses persist in ventral cochlear nucleus long after loss of acoustic input in mice with early-onset progressive hearing loss. *Brain Res.* 1605, 22–30. doi: 10.1016/j.brainres.2015.02.012

Melcher, J. R., Guinan, J. J., Knudson, I. M., and Kiang, N. Y. S. (1996). Generators of the brainstem auditory evoked potential in cat. II. Correlating lesion sites with waveform changes. *Hear. Res.* 93, 28–51. doi: 10.1016/0378-5955(95)00179-4

Morrison, J. H., and Hof, P. R. (1997). Life and death of neurons in the aging brain. *Science* 278, 412–419. doi: 10.1126/science.278.5337.412

Mountain, D. C. (1980). Changes in endolymphatic potential and crossed olivocochlear bundle stimulation alter cochlear mechanics. *Science* 210, 71–72. doi: 10.1126/science.7414321

Murugasu, E., and Russell, I. J. (1996). The effect of efferent stimulation on basilar membrane displacement in the basal turn of the guinea pig cochlea. *J. Neurosci.* 16, 325–332. doi: 10.1523/JNEUROSCI.16-01-00325.1996

Pannese, E. (2011). Morphological changes in nerve cells during normal aging. *Brain Struct. Funct.* 216, 85–89. doi: 10.1007/s00429-011-0308-y

Parthasarathy, T. K. (2001). Aging and contralateral suppression effects on transient evoked otoacoustic emissions. *J. Am. Acad. Audiol.* 12, 80–85.

Radtke-Schuller, S., Seeler, S., and Grothe, B. (2015). Restricted loss of olivocochlear but not vestibular efferent neurons in the senescent gerbil (*Meriones unguiculatus*). *Front. Aging Neurosci.* 7:4. doi: 10.3389/fnagi.2015.00004

Rasmussen, G. L. (1946). The olivary peduncle and other fiber projections of the superior olivary complex. *J. Comp. Neurol.* 84, 141–219. doi: 10.1002/cne.900840204

Reeves, T. M., Phillips, L. L., and Povlishock, J. T. (2005). Myelinated and unmyelinated axons of the corpus callosum differ in vulnerability and functional recovery following traumatic brain injury. *Exp. Neurol.* 196, 126–137. doi: 10.1016/j.expneurol.2005.07.014

Reijntjes, D. O. J., and Pyott, S. J. (2016). The afferent signaling complex: regulation of type I spiral ganglion neuron responses in the auditory periphery. *Hear. Res.* 336, 1–16. doi: 10.1016/j.heares.2016.03.011

Robertson, D., Anderson, C.-J., and Cole, K. S. (1987). Segregation of efferent projections to different turns of the guinea pig cochlea. *Hear. Res.* 25, 69–76. doi: 10.1016/0378-5955(87)90080-3

Rueden, C. T., Schindelin, J., Hiner, M. C., DeZonia, B. E., Walter, A. E., Arena, E. T., et al. (2017). ImageJ2: ImageJ for the next generation of scientific image data. *BMC Bioinform.* 18:529. doi: 10.1186/s12859-017-1934-z

Rutherford, B. R., Brewster, K., Golub, J. S., Kim, A. H., and Roose, S. P. (2018). Sensation and psychiatry: linking age-related hearing loss to late-life depression and cognitive decline. *Am. J. Psychiatry* 175, 215–224. doi: 10.1176/appi.ajp.2017.17040423

Schrode, K. M., Muniak, M. A., Kim, Y.-H., and Lauer, A. M. (2018). Central compensation in auditory brainstem after damaging noise exposure. *eNeuro* 5:ENEURO.250–ENEURO.218. doi: 10.1523/ENEURO.0250-18.2018

Sergeyenko, Y., Lall, K., Liberman, M. C., and Kujawa, S. G. (2013). Age-related cochlear synaptopathy: an early-onset contributor to auditory functional decline. *J. Neurosci.* 33, 13686–13694. doi: 10.1523/JNEUROSCI.1783-13.2013

Shojaei, E., Ashayeri, H., Jafari, Z., Zarrin Dast, M. R., and Kamali, K. (2016). Effect of signal to noise ratio on the speech perception ability of older adults. *Med. J. Islam. Repub. Iran* 30:342.

Simmons, D. D. (2002). Development of the inner ear efferent system across vertebrate species. *J. Neurobiol.* 53, 228–250. doi: 10.1002/neu.10130

Spongr, V. P., Flood, D. G., Frisina, R. D., and Salvi, R. J. (1997). Quantitative measures of hair cell loss in CBA and C57BL/6 mice throughout their life spans. *J. Acoust. Soc. Am.* 101, 3546–3553. doi: 10.1121/1.418315

Statler, K. D., Chamberlain, S. C., Slepecky, N. B., and Smith, R. L. (1990). Development of mature microcystic lesions in the cochlear nuclei of the mongolian gerbil. *Meriones unguiculatus*. *Hear. Res.* 50, 275–288. doi: 10.1016/0378-5955(90)90051-P

Stuart, A., and Phillips, D. P. (1996). Word recognition in continuous and interrupted broadband noise by young normal-hearing, older normal-hearing, and presbyacusic listeners. *Ear Hear.* 17, 478–489. doi: 10.1097/00003446-199612000-00004

Sturrock, R. R. (1987). Changes in the number of neurons in the mesencephalic and motor nuclei of the trigeminal nerve in the ageing mouse brain. *J. Anat.* 151, 15–25.

Suthakar, K., and Ryugo, D. K. (2021). Projections from the ventral nucleus of the lateral lemniscus to the cochlea in the mouse. *J. Comp. Neurol.* 529, 2995–3012. doi: 10.1002/cne.25143

Tang, Y., Nyengaard, J. R., Pakkenberg, B., and Gundersen, H. J. G. (1997). Age-induced white matter changes in the human brain: a stereological investigation. *Neurobiol. Aging* 18, 609–615. doi: 10.1016/S0197-4580(97)00155-3

Terreros, G., Jorratt, P., Aedo, C., Elgoyhen, A. B., and Delano, P. H. (2016). Selective attention to visual stimuli using auditory distractors is altered in alpha-9 nicotinic receptor subunit knock-out mice. *J. Neurosci.* 36, 7198–7209. doi: 10.1523/JNEUROSCI.4031-15.2016

Valero, M. D., Hancock, K. E., Maison, S. F., and Liberman, M. C. (2018). Effects of cochlear synaptopathy on middle-ear muscle reflexes in unanesthetized mice. *Hear. Res.* 363, 109–118. doi: 10.1016/j.heares.2018.03.012

Vetter, D. E., Adams, J. C., and Mugnaini, E. (1991). Chemically distinct rat olivocochlear neurons. *Synapse* 7, 21–43. doi: 10.1002/syn.890070104

Wada, S. I., and Starr, A. (1983). Generation of auditory brain stem responses (ABRs). I. Effects of injection of a local anesthetic (procaine HCI) into the trapezoid body of guinea pigs and cat. *Electroencephalogr. Clin. Neurophysiol.* 56, 326–339. doi: 10.1016/0013-4694(83)90259-6

Warr, W. B., and Guinan, J. J. (1979). Efferent innervation of the organ of corti: two separate systems. *Brain Res.* 173, 152–155. doi: 10.1016/0006-8993(79)91104-1

Winslow, R. L., and Sachs, M. B. (1987). Effect of electrical stimulation of the crossed olivocochlear bundle on auditory nerve response to tones in noise. *J. Neurophysiol.* 57, 1002–1021. doi: 10.1152/jn.1987.57.4.1002

Winslow, R. L., and Sachs, M. B. (1988). Single-tone intensity discrimination based on auditory-nerve rate responses in backgrounds of quiet, noise, and with stimulation of the crossed olivocochlear bundle. *Hear. Res.* 35, 165–189. doi: 10.1016/0378-5955(88)90116-5

Wu, J. S., Yi, E., Manca, M., Javaid, H., Lauer, A. M., and Glowatzki, E. (2020). Sound exposure dynamically induces dopamine synthesis in cholinergic LOC efferents for feedback to auditory nerve fibers. *eLife* 9, 1–27. doi: 10.7554/eLife.52419

Xu, Y., Cheatham, M. A., and Siegel, J. H. (2017). Identifying the origin of effects of contralateral noise on transient evoked Otoacoustic emissions in unanesthetized mice. *J. Assoc. Res. Otolaryngol.* 18, 543–553. doi: 10.1007/s10162-017-0616-x

Yamasoba, T., Lin, F. R., Someya, S., Kashio, A., Sakamoto, T., and Kondo, K. (2013). Current concepts in age-related hearing loss: Epidemiology and mechanistic pathways. *Hear. Res.* 303, 30–38. doi: 10.1016/j.heares.2013.01.021

Zachary, S. P., and Fuchs, P. A. (2015). Re-emergent inhibition of cochlear inner hair cells in a mouse model of hearing loss. *J. Neurosci.* 35, 9701–9706. doi: 10.1523/JNEUROSCI.0879-15.2015

Zettel, M. L., Zhu, X., O'Neill, W. E., and Frisina, R. D. (2007). Age-related decline in Kv3.1b expression in the mouse auditory brainstem correlates with functional deficits in the medial olivocochlear efferent system. *J. Assoc. Res. Otolaryngol.* 8, 280–293. doi: 10.1007/s10162-007-0075-x

Zhu, X., Vasilyeva, O. N., Kim, S., Jacobson, M., Romney, J., Waterman, M. S., et al. (2007). Auditory efferent feedback system deficits precede age-related hearing loss: contralateral suppression of otoacoustic emissions in mice. *J. Comp. Neurol.* 503, 593–604. doi: 10.1002/cne.21402

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Vicencio-Jimenez, Weinberg, Bucci-Mansilla and Lauer. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 04 October 2021
doi: 10.3389/fnins.2021.716689

Behavioral Measures of Cochlear Gain Reduction Depend on Precursor Frequency, Bandwidth, and Level

Kristina DeRoy Milvae†* and *Elizabeth A. Strickland*

Department of Speech, Language, and Hearing Sciences, Purdue University, West Lafayette, IN, United States

Sensory systems adjust to the environment to maintain sensitivity to change. In the auditory system, the medial olivocochlear reflex (MOCR) is a known physiological mechanism capable of such adjustment. The MOCR provides efferent feedback between the brainstem and cochlea, reducing cochlear gain in response to sound. The perceptual effects of the MOCR are not well understood, such as how gain reduction depends on elicitor characteristics in human listeners. Physiological and behavioral data suggest that ipsilateral MOCR tuning is only slightly broader than it is for afferent fibers, and that the fibers feed back to the frequency region of the cochlea that stimulated them. However, some otoacoustic emission (OAE) data suggest that noise is a more effective elicitor than would be consistent with sharp tuning, and that a broad region of the cochlea may be involved in elicitation. If the elicitor is processed in a cochlear channel centered at the signal frequency, the growth of gain reduction with elicitor level would be expected to depend on the frequency content of the elicitor. In the current study, the effects of the frequency content and level of a preceding sound (called a precursor) on signal threshold was examined. The results show that signal threshold increased with increasing precursor level at a shallower slope for a tonal precursor at the signal frequency than for a tonal precursor nearly an octave below the signal frequency. A broadband noise was only slightly more effective than a tone at the signal frequency, with a relatively shallow slope similar to that of the tonal precursor at the signal frequency. Overall, these results suggest that the excitation at the signal cochlear place, regardless of elicitor frequency, determines the magnitude of ipsilateral cochlear gain reduction, and that it increases with elicitor level.

Keywords: cochlear gain reduction, forward masking, medial olivocochlear reflex, elicitor bandwidth, frequency selectivity, psychoacoustics

OPEN ACCESS

Edited by:
David Pérez-González, University of Salamanca, Spain

Reviewed by:
Skyler G. Jennings, The University of Utah, United States
Enrique A. Lopez-Poveda, University of Salamanca, Spain

***Correspondence:**
Kristina DeRoy Milvae
kmilvae@umd.edu

†Present address:
Kristina DeRoy Milvae
Department of Hearing and Speech Sciences, University of Maryland, College Park, College Park, MD, United States

Specialty section:
This article was submitted to Auditory Cognitive Neuroscience, a section of the journal Frontiers in Neuroscience

Received: *29 May 2021*
Accepted: *06 September 2021*
Published: *04 October 2021*

Citation:
DeRoy Milvae K and Strickland EA (2021) Behavioral Measures of Cochlear Gain Reduction Depend on Precursor Frequency, Bandwidth, and Level. Front. Neurosci. 15:716689. doi: 10.3389/fnins.2021.716689

INTRODUCTION

An impressive feat that the human auditory system achieves is the ability to hear sounds that range from low to extremely high intensities. Most neurons in the auditory system respond sensitively to changes over a dynamic range of 30–40 dB, yet we are able to hear over a dynamic range of approximately 120 dB (Viemeister, 1988). This discrepancy between the dynamic range of nerve fibers and the dynamic range of hearing is referred to as the "dynamic range problem" (Evans, 1981;

Viemeister, 1988). One way that the auditory system may overcome the dynamic range problem is by adapting its dynamic range based on the environment. Greater understanding of the adaptive nature of the auditory system has the potential to inform future treatments for hearing loss.

Efferent projections along the entire auditory pathway provide a possible means to adjust the dynamic range. A specific known physiological mechanism that is consistent with this function is the medial olivocochlear reflex (MOCR). The MOCR is an efferent pathway between the brainstem and cochlear outer hair cells that is elicited by sound and acts to decrease cochlear gain, with an onset delay of approximately 25 ms (James et al., 2005; Backus and Guinan, 2006). This gain reduction has been well documented physiologically in neural responses (Winslow and Sachs, 1987; Guinan and Gifford, 1988) and basilar membrane responses (Cooper and Guinan, 2003) in animal models, and in otoacoustic emission (OAE) responses (Backus and Guinan, 2006; Lilaonitkul and Guinan, 2009b) in humans. The MOCR is a bilateral reflex, with evidence suggesting that the ipsilateral pathway, where gain reduction is elicited by preceding sound in the same ear, may be stronger (Lilaonitkul and Guinan, 2012; but see Guinan et al., 2003). This makes the ipsilateral evoked response of interest and the focus of this paper.

The ipsilateral MOCR is elicited by preceding sound, but the frequency of the elicitor affects the magnitude of cochlear gain reduction. Neural measurements in cats have shown that olivocochlear bundle (OCB) fibers have tuning curves that are on average slightly broader than auditory nerve tuning curves and that the feedback loop is frequency-specific, such that preceding sound leads to larger reductions in gain near the cochlear place associated with that frequency (Liberman and Brown, 1986). Bonfils and Puel (1987) examined frequency selectivity of the MOCR by measuring forward masking of compound action potentials (CAPs), the synchronized response of the auditory nerve, in anesthetized guinea pigs to tone pips. These measurements were made with an intact and sectioned crossed (ipsilateral) OCB. Sectioning the crossed OCB caused a decrease in forward masking that occurred when the masker-onset to probe-onset was 40 ms, but not when that same duration was reduced to 30 ms. This suggests efferent contributions to forward masking that occur with a time delay between 30 and 40 ms. Functional tuning curves derived from the decrease in masking were relatively sharp (Q_{10} of 6.6) and centered on the probe frequency, suggesting again that the ipsilateral pathway is elicited in a frequency-specific way and that tuning is similar to that of afferent fibers (Q_{10} of 5–7.3; Bonfils et al., 1986). Similarly, tuning of ipsilateral MOCR effects is sharp when measured with stimulus frequency otoacoustic emissions (SFOAEs) in humans. In SFOAE measurements, the effects of preceding sound may be measured as the combined change in magnitude and phase of the SFOAE, or with magnitude and phase separated. It is not clear what measure is most relevant for the effects of the MOCR on perception. Tuning curves derived from ipsilateral elicitors, with magnitude and phase combined, showed sharp tuning for narrowband or tonal elicitors, with a tip near the probe frequency (Lilaonitkul and Guinan, 2009b). When magnitude and phase were separated, tuning for equal-input elicitors was sharp for magnitude, and more broadly distributed for phase (Lilaonitkul and Guinan, 2012).

In summary, both neural and SFOAE tuning data suggest that ipsilateral elicitation of the MOCR at a cochlear place is primarily driven by energy entering the auditory filter at that cochlear place. However, bandwidth effects have also been measured using SFOAEs that challenge this conclusion. MOCR effects increase with elicitor bandwidth and fixed overall level in a way not explained by additional excitation in the tails of the auditory filter, suggesting that there is integration of elicitation across almost the entire cochlea (Lilaonitkul and Guinan, 2009a) and that broadband noise stimuli are stronger elicitors of cochlear gain reduction than narrowband stimuli (e.g., Guinan, 2018). It is not clear if this bandwidth effect reflects a true difference between the MOCR in human and animal models, or if anesthesia or measurement techniques have led to these differences. Psychoacoustic methods provide an alternative approach to study decreases in cochlear gain in humans which may be due to the MOCR; behavioral measures could provide additional evidence for or against integration of elicitation with wider bandwidths.

Forward masking is a psychoacoustic method to explore cochlear gain reduction with eliciting preceding sound, called a precursor (Krull and Strickland, 2008; Jennings et al., 2009; Roverud and Strickland, 2014; Yasin et al., 2014; DeRoy Milvae and Strickland, 2018). Experimental design can be tailored to the time course of activation of the MOCR to estimate cochlear gain reduction with forward masking (e.g., Yasin et al., 2014; DeRoy Milvae and Strickland, 2018). With this approach (see example paradigm used in this experiment in **Figure 1**), the frequency content of the precursor can be varied to examine how frequency content of the elicitor affects gain reduction. Robust gain

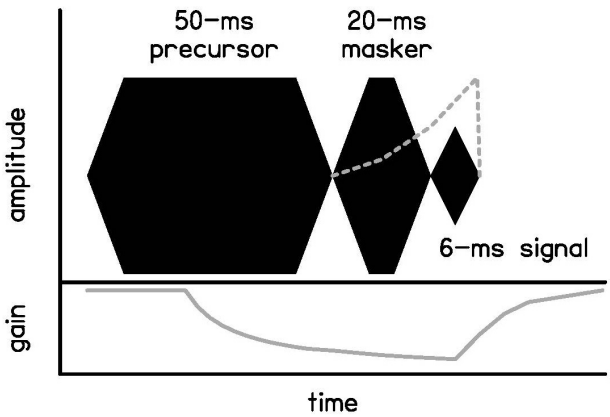

FIGURE 1 | Schematic of the temporal masking paradigm used in this experiment, including a 50-ms precursor, 20-ms masker, and 6-ms signal. The precursor or masker is removed in some experiments, but the temporal relationships are not changed. The frequency content of the precursors and maskers also vary across experiments, but the signal is always presented at 4 kHz. The gray dotted line shows a schematic of the timecourse of forward masking due to neural excitation. The gray solid line shows a schematic of the timecourse of forward masking due to cochlear gain reduction with a precursor present.

reduction has been measured with tonal (Jennings et al., 2009; Roverud and Strickland, 2014) and broadband noise (Yasin et al., 2014; DeRoy Milvae and Strickland, 2018) precursors, but comparisons have not yet been made within-subject to examine if the broadband noise precursors are more effective elicitors of cochlear gain reduction.

However, cochlear gain reduction is not the only mechanism for forward masking. Neural excitation also plays a role in forward masking (see dotted line in **Figure 1**), and models based on this mechanism suggest additivity of masking, meaning that once compression is applied, the intensities of maskers add in their impact on the threshold of a closely following sound (Penner and Shiffrin, 1980; Oxenham and Moore, 1994; Plack et al., 2006). These models assume a static cochlear input-output function, but cochlear gain reduction occurs over time, affecting the cochlear non-linearity (Krull and Strickland, 2008; Roverud and Strickland, 2010). Previous work has shown that models including gain reduction fit data as well or better than those modeled with a static cochlear non-linearity (Jennings and Strickland, 2012; Roverud and Strickland, 2014). In one paradigm with a noise precursor, on-frequency masker, and 4-kHz signal, the signal level was fixed at 15–20 dB SL (sensation level) and masker threshold was measured for a range of precursor levels. The masker level had to be increased to effectively mask the signal with a precursor, more consistent with forward masking due to gain reduction than additivity of masking (Strickland et al., 2018). In this experiment, additivity of masking and gain reduction will again be compared, to establish that the forward masking in this experiment is more consistent with cochlear gain reduction. The paradigm to test this and the predicted results are shown in **Figure 2**. The Power Spectrum Model of masking is used in these predictions, such that detection occurs at a constant effective signal-to-masker ratio at the output of a single auditory filter at the signal frequency [for a review, see Jennings (2021)]. As in a similar paradigm at a lower frequency (DeRoy Milvae and Strickland, 2018), on- and off-frequency maskers will be obtained that elicit the same signal threshold (column 1 of **Figure 2**). An on-frequency precursor will then be added to each condition with the same temporal paradigm shown in **Figure 1**. Predictions are in the second two columns of **Figure 2**, for additivity of masking and gain reduction, respectively. If the additional masking is additive and does not change the cochlear non-linearity, a similar shift in threshold is expected with the addition of the precursor, not dependent on the frequency of the masker (arrows in column 2 of **Figure 2**). However, if the additional masking is related to cochlear gain reduction, no change in threshold is expected with an on-frequency masker, since the signal and masker are on the same function and are equally affected, but a large shift in threshold is expected with an off-frequency masker, since the signal is affected by the gain reduction and the masker is not (Cooper and Guinan, 2006; arrow in column 3 of **Figure 2**).

If the masking associated with the precursor is more consistent with cochlear gain reduction, the effects of precursor frequency content can be explored and interpreted in terms of gain

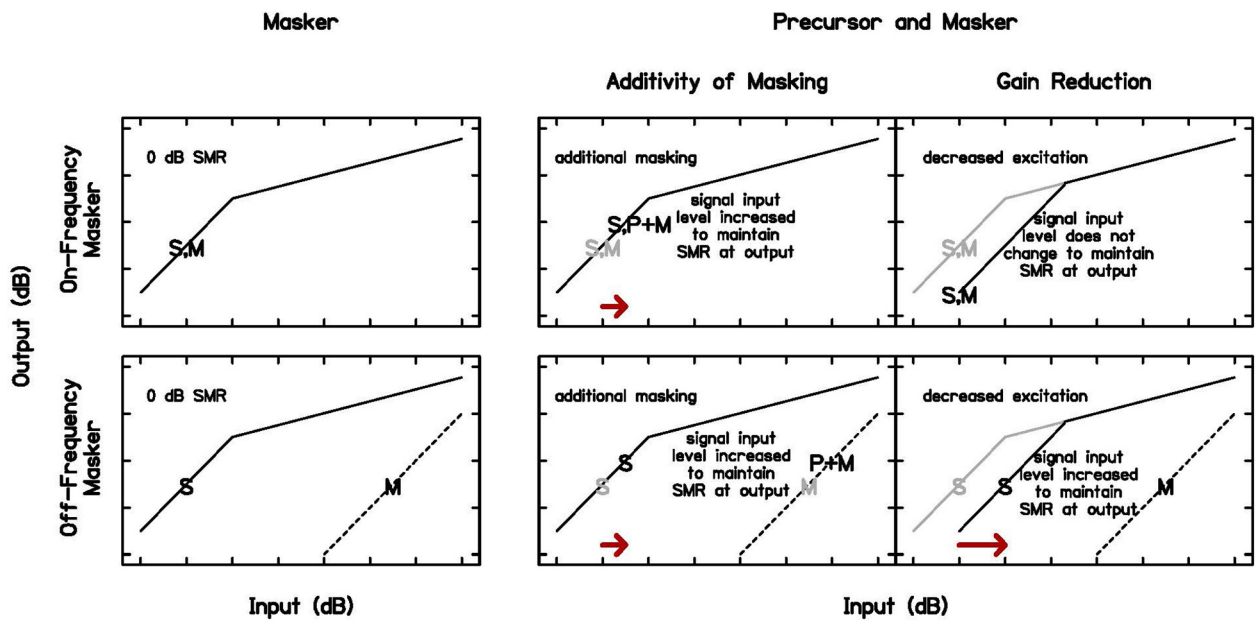

FIGURE 2 | Schematic of cochlear input-output functions and threshold predictions in Experiment 1. Signal threshold (S) occurs at a criterion signal-to-masker ratio (SMR), in this case 0 dB (first column) for an on-frequency (top row) and off-frequency (bottom row) masker (M). With the addition of a precursor (P), predictions differ for forward masking due to additivity of masking or gain reduction. With additivity of masking (second column), a similar shift in signal threshold is expected when the same precursor is presented with equally effective on- and off-frequency maskers (arrows in second column). With gain reduction (third column), a larger shift in signal threshold is expected in the off-frequency case, since the masker is not affected by gain reduction at the signal frequency place (arrow in third column). The input-output functions, S, and M from the first column are repeated in gray in the second and third columns to illustrate the predicted changes with the introduction of a precursor.

reduction. In the case of tonal elicitors, it was hypothesized that gain reduction would occur in a frequency-specific way, as observed with tonal elicitors in previous physiological studies in both animal models and humans (Liberman and Brown, 1986; Bonfils and Puel, 1987; Lilaonitkul and Guinan, 2009b, 2012). In this case, an on-frequency precursor should be a more effective masker than an off-frequency precursor at the same level. Examination of forward masking with increasing precursor level also provides further evidence about tuning; because an on-frequency precursor grows compressively in the auditory channel at the signal frequency place, gain reduction should increase at a slower rate than 1 dB/dB with increasing precursor level. Because an off-frequency precursor should grow linearly in the auditory channel at the signal place, gain reduction should increase at a rate of approximately 1 dB/dB with increasing precursor level. Support for these hypotheses also comes from previous modeling of forward masking data. Modeling off-frequency-elicited gain reduction with level increasing with a slope of 1 dB/dB and on-frequency-elicited gain reduction with level with a shallower slope has predicted forward-masking data well (Roverud and Strickland, 2014). In addition, on- and off-frequency forward masking has been measured previously by Oxenham and Plack (2000), but not interpreted with consideration of cochlear gain reduction.

In the case of broadband noise elicitors, it was hypothesized that they would be more effective elicitors of cochlear gain reduction than tones, as observed in human SFOAE data (Lilaonitkul and Guinan, 2009a). To compare gain reduction with tones and noises, masking at the level of the noise entering an equivalent rectangular bandwidth (ERB; Glasberg and Moore, 1990), an estimated cochlear filter, will be compared to masking at the level of the tonal precursors. Greater masking with the noise would suggest integration across frequency to elicit gain reduction. If, however, the masking with the noise is similar to an on-frequency tone, it would suggest that integration across frequency does not take place and instead that ipsilateral cochlear gain reduction has similar tuning to that seen with afferent nerve fibers.

In this experiment, estimates of cochlear input-output functions were measured for individual participants using a forward masking technique. We hypothesized that shifts in input-output functions with preceding sound at 4 kHz are more consistent with cochlear gain reduction than additivity of masking, as observed previously with a similar paradigm at 1 kHz (DeRoy Milvae and Strickland, 2018). Cochlear gain reduction was examined as a function of the level and frequency content of preceding sound in an effort to examine how the peripheral auditory system remains sensitive across a wide range of input signals, and to examine how elicitation of cochlear gain reduction is tuned. We hypothesized that gain reduction would increase with precursor level, but the slope of increasing gain reduction with increasing precursor level would be shallower with an on-frequency precursor than with an off-frequency precursor, due to cochlear compression of the precursor at the signal place. This would suggest that gain reduction from an ipsilateral elicitor is driven by excitation in an auditory filter at or near the signal frequency, like other forms of forward masking. With a broadband noise precursor, we hypothesized that stronger gain reduction would be elicited than seen with tonal stimuli, as seen with SFOAE measurements in humans. The outcome of this research is an estimate of cochlear gain reduction in decibels, obtained through perceptual measures in humans.

EXPERIMENT 1: FORWARD MASKING WITH A PRECURSOR IS MORE CONSISTENT WITH COCHLEAR GAIN REDUCTION THAN ADDITIVITY OF MASKING

Growth-of-masking (GOM) functions were measured to obtain an estimate of each participant's cochlear input-output function (Oxenham and Plack, 1997; Plack and Oxenham, 1998) with and without preceding stimulation, a precursor (Krull and Strickland, 2008; Jennings et al., 2009; Roverud and Strickland, 2010) under our temporal paradigm (see **Figure 1**). The additional masking with preceding sound could be interpreted as a decrease in cochlear gain, but there are other possible explanations, such as masking due to neural excitation, which predicts additivity of masking given a correction for peripheral compression (Penner and Shiffrin, 1980; Oxenham and Moore, 1994, 1995). A gain reduction hypothesis was tested against additivity of masking using on- and off-frequency forward maskers that resulted in the same signal threshold, making them equally effective maskers of the signal. When the same precursor is added to each condition, additivity of forward masking predicts a similar shift in threshold, regardless of masker frequency. However, gain reduction predicts that the addition of a precursor before an off-frequency masker will lead to a larger shift in threshold (see **Figure 2**). Because an off-frequency masker is processed linearly at the signal place at basal frequencies (Cooper and Guinan, 2006), its gain is not reduced by preceding on-frequency sound, and it is predicted to be a more effective forward masker.

Methods

Participants

Seven young adults (P1–P7) between the ages of 19 and 26 years (median: 21 years) participated in this experiment. All were female except for P5, who was male. All participants had normal audiometric thresholds (15 dB HL or less) at octave frequencies from 0.25 to 8 kHz and present distortion product otoacoustic emissions from 1.5 to 10 kHz. Some participants did not take part in all experiments.

Stimuli

Growth of Masking

Two types of GOM functions were measured for each participant in a forward masking paradigm to estimate the cochlear input-output function at full gain (without reduction in cochlear gain associated with prior sound stimulation) and reduced gain. For the full-gain GOM function, stimuli consisted of a 20-ms, 2.4-kHz tonal masker (including 5-ms $\cos^2$ onset and offset ramps) followed by a 6-ms, 4-kHz tonal signal (including

3-ms $\cos^2$ onset and offset ramps) with no time delay between masker and signal. As in previous studies (e.g., DeRoy Milvae and Strickland, 2018), this masker and signal duration were chosen to be near the estimated onset delay of 20–25 ms for the MOCR (James et al., 2005; Backus and Guinan, 2006), so that there is very little MOCR activation, if any, in this condition. Masker level was fixed between 30 and 95 dB SPL in order to trace out a GOM function for each individual. Signal level was varied to determine the signal level at masking threshold.

A second GOM function at reduced gain was measured for each individual using the same masker and signal, but with the addition of preceding sound before the masker, called a precursor (see **Figure 1** for temporal paradigm). This function was measured with a 50-ms, 40 dB SPL, 4-kHz tonal precursor (including 5-ms $\cos^2$ onset and offset ramps) presented prior to the masker and signal. The precursor duration was 50 ms, as this has been found to be the most effective duration for an on-frequency precursor to shift threshold given this temporal paradigm (Roverud and Strickland, 2014). A level of 40 dB SPL for a tonal on-frequency precursor has been found to produce robust gain reduction in previous studies (Roverud and Strickland, 2010; Jennings and Strickland, 2012).

In addition to the GOM functions, gain reduction was estimated by comparing the signal threshold in quiet to the signal threshold preceded by the precursor and no masker, with a 20-ms silent gap between precursor and signal (in place of the masker). This estimate has shown to be consistent with gain reduction estimates measured with a masker present (DeRoy Milvae and Strickland, 2018; DeRoy Milvae et al., 2021) for listeners with normal thresholds in quiet.

Equally Effective Maskers

On-frequency maskers were identified that were equally effective (produced the same signal threshold) as off-frequency maskers used to measure GOM functions. The 6-ms, 4-kHz signal (including 3-ms $\cos^2$ onset and offset ramps) was fixed at the threshold level obtained when it was preceded by a 20-ms, 2.4-kHz masker (including 5-ms $\cos^2$ onset and offset ramps). The level of a 20-ms, 4-kHz masker (including 5-ms $\cos^2$ onset and offset ramps) was then varied to measure threshold and find the lowest masker level where the signal could be detected. This level was then confirmed to produce the same signal threshold as the off-frequency masker by fixing the masker level and varying the signal level. This was done for two points on the lower leg of the GOM function for each participant, although an effect of masker frequency with the addition of a precursor was expected as long as the point chosen was not affected by compression.

To examine whether shifts in forward masking with a precursor were more consistent with gain reduction than additivity of masking, an identical precursor was presented before the two equally effective maskers and signal threshold was measured in each condition (measurements from the GOM function used for the off-frequency conditions). Additivity of masking predicts that adding a 50-ms, 40 dB SPL, 4-kHz precursor (including 5-ms $\cos^2$ onset and offset ramps) before the on-frequency masker and off-frequency masker that produce the same signal threshold should cause an identical shift in threshold (see column 2 of **Figure 2**). This method does not rely on the measurement of the input-output function for interpretation. It was hypothesized that a larger shift in signal threshold would be seen for the off-frequency condition, more consistent with precursor masking related to cochlear gain reduction (see column 3 of **Figure 2**).

Procedure

The experiment took place in a double-walled sound-attenuating booth (IAC, Bronx, NY, United States). Tucker–Davis Technologies (TDT, Alachua, FL, United States) hardware was used. Stimuli were digitally generated at a sampling rate of 25 kHz. They were then sent to four separate digital-to-analog channels (TDT DA3-4, 16-bit), low pass filtered at 10 kHz (TDT FT5 and FT6-2), mixed (TDT SM3), buffered (TDT HB6), and output to the participant's right ear *via* an ER-2 (Etymotic Research, Inc., Elk Grove Village, IL, United States) insert earphone. This insert earphone has a flat frequency response at the eardrum for frequencies from 0.25 to 8 kHz. Participants wore both the left and right earphones, even though sound was not presented to the left ear, to reduce interference from ambient noise.

Participants performed a three-interval forced choice task. Intervals were separated by 500 ms of silence and participants indicated the interval containing the signal by pressing a key. Visual indicators were used to identify the intervals and feedback was given to indicate the veracity of the participant's choice. The signal level was adjusted while the masker level was held constant to approximate a detection threshold of 70.7% correct on the psychometric function (Levitt, 1971) for the range of masker levels tested. To determine the on-frequency masker levels needed to elicit a similar signal threshold as off-frequency maskers, the masker level was adjusted while the signal level was held constant. Participants completed 2–5 h of training on GOM tasks to control for learning effects and 1–3 h of training with on-frequency masker conditions for the equally effective maskers task. Less training was needed on this task because participants were already familiar with the general forward masking task. Two runs for each condition collected on the final day of participation are included in the experimental data. However, on-frequency masked thresholds of P2 continued to show high variability after training. For this reason, more than two estimates of each threshold were attempted for this participant, with an average of 3.5 threshold estimates measured per condition that did not have to be removed from experimental data due to high standard deviations. Off-frequency conditions were also repeated for this subject instead of using the measurements from the GOM function, so that measurements with equally effective maskers were collected at a similar point in time for this highly variable listener. In addition, an experimenter error led to three thresholds collected for P5 in the 65 dB SPL off-frequency masker and precursor condition (reduced gain GOM function and off-frequency equally effective masker condition with a precursor), but this additional threshold was similar to the first two measured and was not believed to influence the results.

During each masked trial, high pass noise was presented to limit off-frequency listening (Nelson et al., 2001). It began 50 ms before the first stimulus and ended 50 ms after the signal. The noise was presented at a spectrum level 50 dB below the signal level (varying adaptively with the signal level), and had 5-ms $\cos^2$ onset and offset ramps and a bandwidth of 4.8–8.0 kHz. Because P2 demonstrated difficulty with the tasks when the high pass noise was present, resulting in inconsistent thresholds across trials, the noise was removed during testing for this participant.

Each run consisted of 50 trials. The step size was 5 dB before the second reversal in signal (or masker) level, and then the step size decreased to 2 dB. Runs were excluded if the standard deviation was greater than 5 dB for one or two final runs or if less than six reversals were present. The final even number of reversals at the 2-dB step size were averaged to estimate threshold for each run.

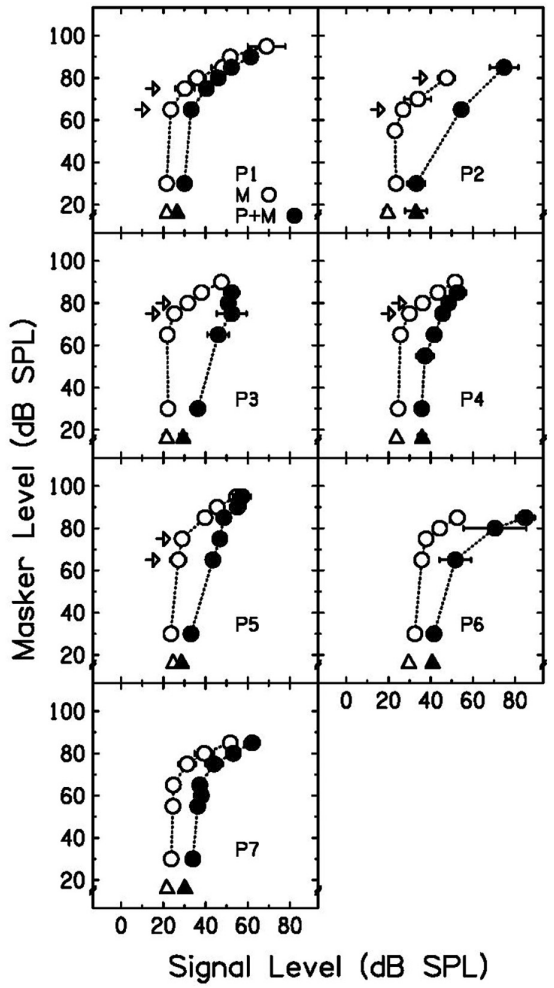

FIGURE 3 | Individual GOM functions and masker-absent gain reduction estimates. Signal thresholds for the masker-alone condition are plotted as open circles and signal thresholds with the addition of a precursor are plotted as filled circles. Signal threshold without a preceding masker is plotted as an open triangle, and signal threshold with a precursor and 20-ms delay is plotted as a filled triangle. The difference between the triangles is the masker-absent gain reduction estimate. Arrows indicate the off-frequency masker levels used in the equally-effective-masker conditions. Error bars represent one standard deviation.

Results

Growth of Masking

Growth-of-masking functions without a precursor (open circles) and with a precursor (filled circles) are plotted in **Figure 3**. Open triangles represent the signal threshold when the signal is presented alone. Filled triangles represent the signal threshold when the precursor is present but there is no masker (20-ms gap of silence between precursor and signal). As shown in previous work (Krull and Strickland, 2008; Jennings et al., 2009; Roverud and Strickland, 2010), the precursor shifted the lower leg of the GOM function to higher signal levels (a rightward shift). This shift is consistent with a decrease in cochlear gain. P2 had a limited number of thresholds for the precursor condition because this participant's runs often resulted in standard deviations that were above 5 dB, and those thresholds were not included. It was observed that the masker-absent gain reduction estimate (difference between open and filled triangles in **Figure 3**) was a reasonable estimate for the gain reduction observed by the shift in the GOM function, as shown previously (DeRoy Milvae and Strickland, 2018; DeRoy Milvae et al., 2021).

Equally Effective Maskers

Individual signal thresholds are shown in **Table 1** and average threshold shifts with the addition of a precursor at two masker frequencies are shown in **Figure 4**. As was shown in **Figure 3**, the precursor shifted signal thresholds to higher levels when the masker was 2.4 kHz. In the 4 kHz masker case, there was a much smaller shift in threshold. One-tailed *t*-tests (with a Holm-Bonferroni correction) were performed to test for significance that the threshold for the off-frequency condition with an added precursor was higher than that of the on-frequency condition with an added precursor at the individual level, and significant differences ($p < 0.05$) are noted by asterisks in **Table 1**. P1, P2, and P3 showed a significantly higher threshold for the off-frequency condition with an added precursor at one level of matched threshold, $t(2) = 8.03$, $p = 0.046$; $t(6) = 8.43$, $p < 0.001$; and $t(2) = 5.23$, $p = 0.041$; respectively. P5 showed this same effect at two levels of matched threshold. For a matched threshold of 27 dB SPL, $t(3) = 6.98$, $p = 0.027$, and for a matched threshold of 29 dB SPL, $t(2) = 9.57$, $p = 0.043$. Other participants showed a similar trend that did not reach significance. In addition, the data were averaged across participants by taking the average difference between the precursor condition and the masker-alone condition for each masker frequency (averaging the two levels for each participant). A one-tailed *t*-test was performed for these data and there was a significant difference between the average change in threshold for a 2.4-kHz masker and a 4-kHz masker when an identical precursor is added, $t(8) = 4.91$, $p = 0.006$.

Discussion

The shift in signal threshold with a precursor and no masker (difference between open and filled triangles in **Figure 3**) was demonstrated to be a reasonable estimate of gain reduction,

TABLE 1 | Individual data with equally effective maskers.

Participant	Masker level		Masked signal threshold		Masked signal threshold with 4-kHz precursor	
	2.4-kHz masker	4-kHz masker	2.4-kHz masker	4-kHz masker	2.4-kHz masker	4-kHz masker
P1	65	16	23.53 (1.60)	22.97 (0.20)	*33.17 (0.24)	27.67 (0.94)
	75	33	30.20 (4.53)	29.98 (1.21)	40.42 (3.24)	35.85 (0.68)
P2	65	21	28.65 (1.09)	33.85 (11.54)	*50.80 (2.90)	36.42 (1.89)
	80	45	45.41 (4.59)	51.68 (5.56)	68.36 (9.77)	57.89 (2.31)
P3	75	14	25.25 (1.06)	22.30 (0.42)	52.40 (7.07)	26.27 (0.09)
	80	30	31.61 (1.15)	29.72 (0.08)	*50.77 (2.27)	35.16 (0.79)
P4	75	28	29.93 (1.52)	30.76 (1.28)	45.60 (1.98)	39.24 (4.38)
	80	37	36.18 (1.44)	41.81 (3.37)	48.46 (0.94)	43.93 (3.21)
P5	65	21	27.12 (3.84)	26.08 (4.36)	*43.48 (2.17)	32.19 (0.04)
	75	27	28.89 (1.97)	29.78 (1.96)	*46.78 (1.96)	33.54 (0.05)

Mean masked signal threshold (dB SPL) is shown for two masker levels chosen to elicit similar masked thresholds with both masker frequencies.
Masked signal threshold with the addition of the same precursor is shown for off- and on-frequency masker conditions.
One standard deviation is shown in parentheses.
Signal thresholds for off-frequency masker conditions with a precursor that were significantly higher than the corresponding on-frequency masker condition with a precursor are indicated with an asterisk.

as observed previously (Roverud and Strickland, 2010; DeRoy Milvae and Strickland, 2018). However, in some cases it was lower than that observed in the GOM function; for example, the masker-absent threshold shift was smaller than that with a masker present for P3. Lower estimates may be found with this method since the MOCR can reduce the spontaneous rate of auditory nerve fibers (Guinan and Gifford, 1988). Therefore, the masker-absent estimate of gain reduction may sometimes underestimate gain reduction.

With equally effective maskers that differed in frequency, a larger shift in threshold was induced for the off-frequency masker condition than for the on-frequency masker condition with the introduction of an on-frequency precursor (**Figure 4**). Since the change in threshold depended on masker frequency, the masking provided by the precursor was more consistent with gain reduction than neural excitation alone. Additivity of masking would predict a similar change in threshold, regardless of masker frequency. The current data show that when the effects of a precursor on an on-frequency and off-frequency masking condition are compared, the change in signal threshold is not easily explained by additivity of masking. This difference in threshold shift measured was consistent with gain reduction in that the precursor in both cases elicits gain reduction at the 4-kHz place, differentially affecting the on- and off-frequency maskers. Since the 2.4-kHz masker should have an approximately linear response at the 4-kHz place, it is not affected by the gain reduction elicited by the precursor and is thus a more effective masker than the 4-kHz masker in this condition. This leads to a greater shift in threshold for the off-frequency masker condition. Even with this differential effect, some change in threshold can be seen for the on-frequency masker. This effect is still consistent with gain reduction. It can occur if the gain is decreased enough that the signal becomes inaudible. Alternatively, residual additivity of masking, after accounting for gain reduction, could also explain the increase in threshold with an on-frequency masker.

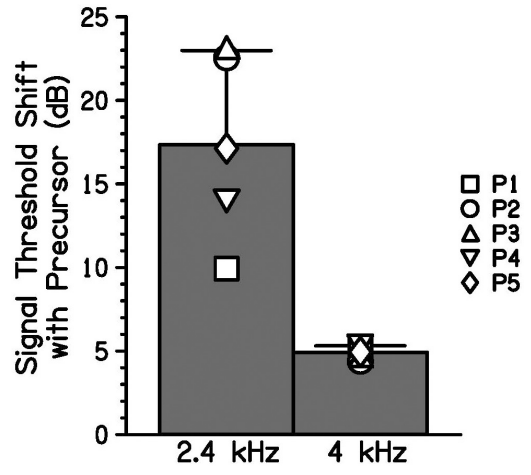

FIGURE 4 | Bars indicate the group average increase in signal threshold with a precursor preceding equally effective off-frequency (2.4-kHz) and on-frequency (4-kHz) maskers. Signal threshold shift with a precursor was averaged for two matched signal levels for each participant (symbols). Error bars represent one standard deviation.

This result is similar to that observed previously at 4 kHz (Jennings et al., 2009) and at a lower signal frequency (DeRoy Milvae and Strickland, 2018). A differential effect of a precursor on masking of a signal by on- and off-frequency maskers below the signal frequency has also been seen in studies in which the signal level was fixed and the masker level was varied to measure a psychoacoustic tuning curve or a temporal masking curve. In these cases, the addition of the precursor decreases the masker level needed to mask the signal for the off-frequency masker, but not for the on-frequency one. This has been seen with a contralateral precursor (Kawase et al., 2000; Fletcher et al., 2016) and an ipsilateral precursor (Jennings and Strickland, 2012).

Additional evidence supporting a gain reduction explanation comes from Roverud and Strickland (2014), a study exploring differences in forward masking with on- and off-frequency precursors. They measured the shift in threshold following an off-frequency masker produced by an on- or off-frequency precursor, as a function of precursor duration. For the 2.4-kHz precursor, threshold increased with precursor duration for

durations up to 160 ms. For the 4-kHz precursor, however, threshold increased with precursor duration up to 50 ms, but then either plateaued or in some cases oscillated. This was modeled using a temporal window model combined with gain reduction elicited by the precursor. For an on-frequency precursor, the precursor itself was affected by gain reduction, and thus effectiveness fluctuated with duration. The off-frequency precursor was not affected by gain reduction within the signal channel, and thus effectiveness continued to grow with duration.

EXPERIMENT 2: SIGNAL THRESHOLD WITH INCREASING LEVEL OF TONAL PRECURSORS

The results of Experiment 1 support the theory that a shift in signal threshold with a precursor reflects gain reduction. In that case, it is of interest to examine gain reduction as a function of precursor frequency and level, to examine the tuning of cochlear gain reduction elicitation. The results of previous studies suggest that gain reduction may increase at a slope of approximately 1 dB/dB of increasing precursor level for a masker well below the signal frequency (Oxenham and Plack, 2000; Roverud and Strickland, 2014), and increase at a shallower slope for a masker at the signal frequency (Plack and Oxenham, 1998; Oxenham and Plack, 2000; Roverud and Strickland, 2014). This experiment replicates and builds on aspects of the design of Oxenham and Plack (2000), and results are interpreted taking into account a gain reduction hypothesis.

Methods

Participants

The same seven participants from Experiment 1 (P1–P7) between the ages of 19 and 26 years (median: 21 years) participated in this experiment.

Stimuli

A 50-ms, 2.4- or 4-kHz precursor (including 5-ms $\cos^2$ onset and offset ramps) was presented with a 20-ms silent gap before the signal, a 6-ms, 4-kHz tone (including 3-ms $\cos^2$ onset and offset ramps). Precursor levels were fixed between 20 and 90 dB SPL for the 4-kHz precursor (on-frequency) condition and between 60 and 95 dB SPL for the 2.4-kHz (off-frequency) condition in order to trace out changes in signal threshold for each individual.

Procedure

Equipment and procedures were identical to those of Experiment 1 with the following exceptions. Signal level was adjusted while the precursor level was held constant during measurements (the signal level was not held constant for any measurements in this experiment). Additionally, approximately 2–5 h of training were completed on these tasks to control for learning effects.

Results

Gain reduction was estimated by subtracting each participant's threshold for the signal alone (quiet threshold, shown as open triangles in **Figure 3**) from their signal threshold with the precursor measured in this experiment. A shift in signal threshold due to the presence of the precursor was interpreted as estimated gain reduction. Estimated gain reduction is plotted as a function of precursor level in **Figure 5**. Gain reduction increased with precursor level more rapidly in the off-frequency masker condition (pink diamonds) than in the on-frequency masker condition (green circles).

To test whether the slope was significantly different with precursor frequency, a linear mixed-effects model (LMM) was used to fit this data set. All data points were included that provided over 4 dB of estimated gain reduction to remove floor effects that would affect the slope (the data shown in **Figure 5**). The dependent variable was estimated gain reduction (dB). Fixed effects included in the model were precursor frequency (a categorical variable of 2.4 or 4 kHz, with 4 kHz chosen as the reference level) and precursor level (a continuous variable), and the interaction between precursor frequency and level. A significant interaction would be interpreted as a significant difference in slope of the functions with precursor frequency. Random intercepts and slopes were included as random effects in the maximal model for precursor frequency by participant. Model testing was completed using R 4.0.3 (R Core Team, 2020) and the "buildmer" version 1.5 (Voeten, 2020) package. The buildmer function ordered effects using the likelihood-ratio test statistic (LRT), followed by backward-elimination model testing based on the significance of changes in log-likelihood. With this approach and the maximal model as input (Barr et al., 2013), a model was

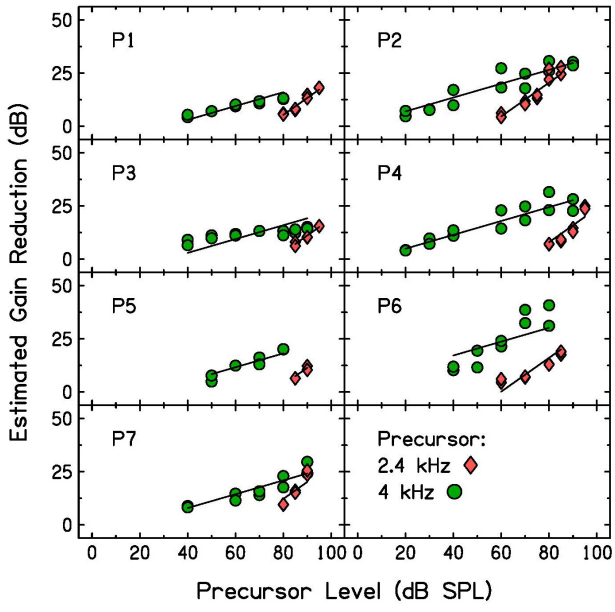

FIGURE 5 | Individual estimated gain reduction with precursor level measured with 2.4-kHz (pink diamonds) and 4-kHz (green circles) precursors. Estimated gain reduction was calculated by subtracting quiet threshold for the signal from the signal threshold for each condition. The two data points measured for each precursor level and included in the analysis are plotted. Linear mixed-effects model fits (see **Table 2** for model summary) are plotted as lines over the data.

found that both converged and best fit the data (Matuschek et al., 2017; Voeten, 2020). The Satterthwaite approximation (Luke, 2017) was used to generate *p*-values. The model summary for the model of best fit is in **Table 2**. There is a significant interaction between precursor level and frequency, such that the slope is significantly shallower in the on-frequency precursor condition. The slope of the on-frequency precursor condition fit was 0.33, and the slope of the off-frequency precursor condition fit was 0.47 higher, with a slope of 0.80 (see **Table 2**). Thus, the slope was far shallower in the on-frequency precursor condition.

Discussion

The slope of estimated gain reduction with increasing precursor level depended on precursor frequency (**Figure 5** and **Table 2**). This finding is consistent with previous model assumptions. Roverud and Strickland (2014) used a gain reduction model based on the timecourse of the MOCR, followed by a temporal window model, to model forward masking with increased precursor duration. In the model, the input to the gain-reduction module was assumed to grow compressively for an on-frequency precursor, using the compression derived from each listener's input-output function. For the off-frequency precursor, the input to the gain reduction module was assumed to grow linearly. With these assumptions, the data were fit well. The model incorporating gain reduction fit the data better than a model using only a temporal window. The results of the present study are consistent with those results, in that the growth of gain reduction with precursor level has a shallow slope with an on-frequency precursor, and has a more linear slope with an off-frequency precursor. This suggests physiologically that the precursor is processed at the signal place, such that on-frequency sound is affected by cochlear compression, and that the output of the cochlear non-linearity serves as a local input to cochlear gain reduction.

The estimated gain reduction measured in this experiment is very similar to forward masking measured by Oxenham and Plack (2000). Oxenham and Plack (2000) measured growth of forward masking with masker level for 0-, 10-, and 30-ms delays between a 200-ms masker (on- and off-frequency) and a 10-ms, 4-kHz signal. For on-frequency maskers, the slope of increased masking with increased masker level became shallower with longer delays between the masker and signal. For off-frequency maskers, delay did not affect the slope. Although Oxenham and Plack (2000) did not interpret their data in terms of cochlear gain reduction, gain reduction estimates can be made from the 30-ms delay data they presented by subtracting quiet threshold for each participant from the masked thresholds presented. On-frequency data show approximately 20–35 dB of maximum gain reduction for the participants and off-frequency data shows approximately 25–40 dB of maximum gain reduction for the participants. The present results had similar ranges of maximum gain reduction. Although Oxenham and Plack (2000) interpreted their results using a temporal window model (additivity of masking), this model did not predict the rollover (decrease in threshold) observed with increased masker duration. Similar rollover effects have been modeled well with gain reduction (Roverud and Strickland, 2014). Therefore, cochlear gain reduction is an alternative explanation for the Oxenham and Plack (2000) data.

The on- and off-frequency functions are consistent with the idea that excitation at the signal place, regardless of the frequency presented to the ear, leads to cochlear gain reduction which increases with level. Gain reduction increases at a slower rate if the precursor itself is compressed in the system, as is the case with an on-frequency precursor.

TABLE 2 | LMM summary describing the effects of precursor level and frequency on gain reduction estimates (dB).

	Estimated gain reduction (dB)			
Fixed effects	**Estimate**	***SE***	***t***	***p***
Intercept	−4.32	2.49	−1.74	0.11
Level	**0.33**	**0.02**	**16.03**	**<0.001**
Frequency (2.4 > 4 kHz)	**−49.47**	**6.27**	**−7.89**	**<0.001**
Level × Frequency	**0.47**	**0.07**	**6.37**	**<0.001**

Random effects	**Variance**	***SD***	**Correlations**
By-participant intercepts	31.23	5.59	
By-participant frequency slopes	16.18	4.02	−0.02
Residual	11.98	3.46	

Significant fixed effects are shown in bold.
The p-values were calculated with a Satterthwaite approximation.

EXPERIMENT 3: SIGNAL THRESHOLD WITH INCREASING LEVEL OF BROADBAND NOISE PRECURSORS

There is evidence from otoacoustic emission data that stimuli with wider bandwidths are more effective elicitors of ipsilateral cochlear gain reduction (Lilaonitkul and Guinan, 2009b). It is believed that broadband noise is more efficient in eliciting the MOCR than stimuli with smaller bandwidths, such as narrowband noise or tones (Norman and Thornton, 1993; Maison et al., 2000; Guinan et al., 2003). In addition, many studies investigating contralateral MOCR activity in humans with otoacoustic emissions use 60-dB SPL broadband noise as an elicitor (e.g., Guinan et al., 2003; Backus and Guinan, 2006; Francis and Guinan, 2010). However, in psychoacoustic studies, both tones and noises have been used as elicitors, and have shown large amounts of gain reduction (Jennings et al., 2009; Roverud and Strickland, 2014; Yasin et al., 2014; DeRoy Milvae and Strickland, 2018). Therefore, it is also of interest to measure growth of gain reduction when the precursor is a broadband noise and compare to growth of gain reduction with pure tones, to examine if behavioral measures of cochlear gain reduction with a wideband stimulus are consistent with integration of gain reduction elicitation across the cochlea.

Methods

Participants

P1–P5 from Experiments 1 and 2 returned to complete Experiment 3. However, P5 was removed from the study for the

reasons described in the "Stimuli" section. The ages of P1–P4 range 20–26 years (median: 22 years).

Stimuli

Gain reduction was estimated as a function of precursor level, as in Experiment 2, but the precursor in this experiment was a broadband noise (BBN) rather than a pure tone. The BBN was 80–10,000 Hz wide and levels ranged from 24 to 64 dB overall RMS level. Higher levels were not tested to avoid confounding effects from the middle-ear muscle reflex, elicited at a lower level by noises than tones (e.g., Schairer et al., 2007). The BBN precursor was 50-ms in duration (including 5-ms $\cos^2$ onset and offset ramps) and was presented with a 20-ms gap before the 6-ms, 4-kHz (including 3-ms $\cos^2$ onset and offset ramps) signal. It was hypothesized that the increase in estimated gain reduction with precursor level would have a steeper slope than that of an on-frequency precursor if elicitation is integrated across cochlear place. Alternatively, if on-frequency energy of the noise dominates elicitation of gain reduction at the signal-frequency place, a similar slope was expected as seen with on-frequency tones.

In the MATLAB program, an error was present which led to the removal of data from P5. When generating the BBN, the program used a frozen noise for the two intervals without the signal and generated a second noise for the interval with the signal. Because of this, participants were able to listen for a change in the noise (which was not changing in level across presentations) rather than listen for the signal. P5 was the only participant to use this cue, leading to impossibly low thresholds. Because other participants' thresholds were not impossibly low, it was assumed that they did not use the noise cue available and their data are presented, although some contribution of this cue cannot be entirely ruled out.

Procedure

Experiments took place in a sound-attenuated booth (IAC, Bronx, NY, United States). A custom program developed using MATLAB software (2011a, The Math Works, Natick, MA, United States) was used to present stimuli. Stimuli were generated in MATLAB *via* a Lynx TWO-B sound card (Lynx Studio Technology, Inc., Costa Mesa, CA, United States). They were then buffered (TDT HB6) and presented to a right ER-2 insert earphone.

The procedure was identical to that of Experiment 2 except for a difference in the criterion for number of trials in the adaptive procedure. The step size was 5 dB before the fourth reversal; it then decreased to a step size of 2 dB for the remainder of the trials. Trials continued until 12 reversals were completed, and the final eight reversals were averaged to establish threshold.

Results

Estimated gain reduction with BBN precursors is plotted as a function of precursor level in **Figure 6** (purple hourglass symbols), with the tonal precursor data from Experiment 2. Gain reduction estimates were calculated in the same way as in Experiment 2; quiet threshold for the signal alone was subtracted from the threshold for the signal with the BBN precursor at each level. Level of the BBN is plotted as the decibel level per equivalent rectangular bandwidth (dB/ERB) of the noise. This calculation was done to approximate the level of the sound entering the cochlear filter centered at the signal frequency, 4 kHz. The equation used to calculate the ERB (Glasberg and Moore, 1990) is shown below, where F is the frequency of the signal in kHz and ERB is the filter bandwidth at that frequency.

$$ERB = 24.7\,(4.37F + 1) \tag{1}$$

The ERB at 4 kHz is 456.46 Hz with this calculation. Next, to find the decibel level entering the filter (dB/ERB), the following equation was used, where SL is the spectrum level of the BBN.

$$\frac{dB}{ERB} = SL + 10\log(ERB) \tag{2}$$

This transformation was done to make the units more comparable to those used to describe the tonal precursors; the level given is an estimate of the energy within a critical band at the signal place. Note that the off-frequency tone at 2.4 kHz does not fall within this filter; excitation from the off-frequency precursor would fall in the tail of the auditory filter, which is not captured using ERB.

Participants 1, 3, and 4 show similar estimates of gain reduction for both 4-kHz and BBN precursors when the noise level is plotted in dB/ERB units. This suggests that energy in the critical band filter dominates the elicitation of gain reduction, and that tones and noise are each able to reduce gain for a signal. P2, however, does not show similar gain reduction estimates for both 4-kHz and BBN precursors. P2's BBN function has steeper growth than that measured with a 4-kHz tone. P2 is the same participant who did not tolerate the high pass noise presented when measuring GOM functions, indicating possible difficulty with listening tasks.

A LMM was again used to test for significant differences in slope with precursor frequency content. The statistical software

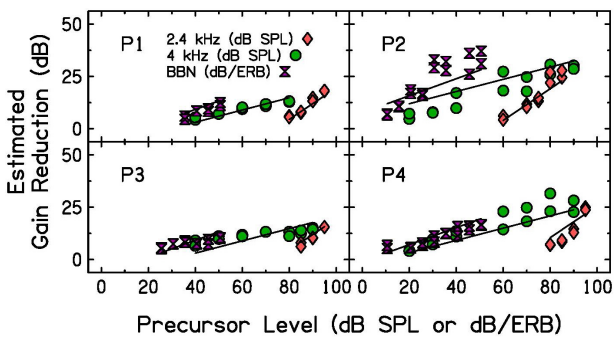

FIGURE 6 | Individual estimated gain reduction with precursor level measured with BBN (purple hourglasses), 2.4-kHz (pink diamonds, identical to data presented in **Figure 5**) and 4-kHz (green circles, identical to data presented in **Figure 5**) precursors. Precursor level is plotted in dB per equivalent rectangular bandwidth (ERB) for the noise precursors. Estimated gain reduction is calculated by subtracting quiet threshold for the signal from the signal threshold for each condition. LMM fits (see **Table 3** for model summary) are plotted as lines over the data.

and packages, criteria for included data points (data shown in **Figure 6** included), dependent variable, fixed effects, reference levels, and random effects were the same as described for the analysis of Experiment 2, with the following exception. The fixed effect of precursor frequency was a categorical variable with three levels rather than two: 2.4 kHz, 4 kHz, and BBN. The model summary for the model of best fit is in **Table 3**. There was a significant interaction between precursor level and frequency content, such that the slope was significantly steeper in the off-frequency precursor condition compared to the on-frequency precursor condition ($p < 0.001$), as shown in Experiment 2, and the slope is significantly steeper with a BBN precursor compared to a 4-kHz precursor ($p = 0.03$). The slope of the on-frequency precursor condition fit was 0.29, the slope of the off-frequency precursor condition fit was approximately 0.46 higher, with a slope of 0.76, and the slope of the broadband noise precursor condition was approximately 0.11 higher, with a slope of 0.41 (see **Table 3**). Thus, the slope was far shallower in the on-frequency and broadband noise precursor conditions than in the off-frequency condition.

Discussion

As precursor level increased, gain reduction increased with increasing precursor level at a rate less than 1 dB/dB for a BBN precursor. Estimates of gain reduction with BBN precursors increased at a significantly shallower rate than off-frequency precursors measured in Experiment 2 and at a significantly steeper rate than on-frequency precursors measured in Experiment 2. The gain reduction estimates with BBN and on-frequency precursors were closer in slope than the gain reduction estimates with BBN and off-frequency precursors (**Figure 6**). It was concluded that the shallow slope for both BBN and on-frequency precursors was likely related to cochlear compression.

The similarity in this study between the magnitude of gain reduction elicited by BBN and on-frequency precursors was surprising, since OAE data have suggested that noises are more robust elicitors of the ipsilateral MOCR (Lilaonitkul and Guinan, 2009a). For three of the four subjects tested, the data points for a BBN precursor are very consistent with those obtained with on-frequency tonal precursors. P2 showed a different pattern, but also had overall difficulty with the listening tasks (inferred based on inconsistency of threshold measurements). Because of this difficulty, high pass noise was removed when GOM functions were measured for P2 to obtain more consistent thresholds. The lack of consistency argues that this listener had more trouble with the task, rather than broadband noise being a stronger elicitor of gain reduction.

The estimated gain reduction measured with a BBN precursor in this study can be compared to that of other psychoacoustic studies. Yasin et al. (2014) used a different forward masking technique, in which signal and masker durations were adjusted within a 25-ms masker-signal complex to estimate the input-output function, and a precursor was presented before the masker at delays of 0-, 50-, 100-, and 200-ms. The masker was either on- or off-frequency, and the precursor was an on-frequency narrowband noise. A comparison was made between on- and off-frequency masker data to estimate cochlear gain. With this approach, they found a similar increase in cochlear gain reduction with precursor level; they reported a slope of 0.33 for the 0-ms delay condition, which is similar to that measured in this experiment for on-frequency and BBN precursors. Maximum gain reduction was approximately 25 dB (Yasin et al., 2014), consistent with the current results. In another study using pink-noise precursors, approximately 10 dB of gain reduction was estimated with a 60 dB SPL overall precursor level with a 50-ms duration, again consistent with the present results with a BBN (DeRoy Milvae and Strickland, 2018). The estimated gain reduction in this study and Yasin et al. (2014) is largely consistent with physiological measures (Russell and Murugasu, 1997; Cooper and Guinan, 2006). Maximum gain reduction of 25 dB is larger than the 15–20 dB of maximum gain reduction measured physiologically (Russell and Murugasu, 1997; Cooper and Guinan, 2006). A difference between the psychoacoustic measures and physiological measure is that cochlear gain reduction was evoked by sound psychoacoustically, but by electrical pulses physiologically (Dolan et al., 1997; Russell and Murugasu, 1997). Therefore, the psychoacoustic estimates of gain reduction confirm that the decibel levels measured with electrical stimulation are plausible in a natural listening situation. It is possible that greater gain reduction emerges psychoacoustically due to differences in the stimulation mode, differences across species, or forward masking contributions unrelated to cochlear gain reduction at high levels.

TABLE 3 | LMM summary describing the effects of precursor level and frequency content on gain reduction estimates (dB).

	Estimated gain reduction (dB)			
Fixed effects	**Estimate**	***SE***	***t***	***p***
Intercept	−3.40	3.86	−0.88	0.42
Level	**0.29**	**0.03**	**11.38**	**<0.001**
Frequency (2.4 > 4 kHz)	**−47.50**	**6.95**	**−6.83**	**<0.001**
Frequency (BBN > 4 KHz)	1.52	2.33	0.65	0.52
Level × Frequency (2.4 > 4 kHz)	**0.46**	**0.08**	**5.53**	**<0.001**
Level × Frequency (BBN > 4 kHz)	**0.11**	**0.05**	**2.17**	**0.03**

Random effects	**Variance**	***SD***
By-participant intercepts	48.90	6.99
Residual	15.62	3.95

Significant fixed effects are shown in bold.
The p-values were calculated with a Satterthwaite approximation.

GENERAL DISCUSSION

In Experiment 1, GOM functions were measured with and without preceding sound to obtain an estimate of each participant's cochlear input-output function at full gain and with decreased gain (**Figure 3**). In addition, the theory that the masking provided by the precursor is due to decreased cochlear gain was tested. Equally effective on- and off-frequency maskers were found, and the same precursor was

added to each condition. Additivity of forward masking predicts that this addition of the precursors would lead to a similar shift in signal threshold, regardless of the masker frequency. However, there was a larger shift in signal threshold for the off-frequency masker condition (**Figure 4**). This is consistent with a gain reduction hypothesis to explain the additional forward masking. In Experiments 2 and 3, gain reduction was estimated for a range of precursor levels (**Figures 5**, **6**). On-frequency, off-frequency, and BBN precursors were used. Increases in estimated gain reduction with increased precursor level varied in slope, with a shallower slope for on-frequency and BBN precursors than off-frequency precursors (**Tables 2**, **3**). It is possible that the shallow slopes seen with the on-frequency and BBN precursors were due to cochlear compression, and that elicitation of gain reduction with level reflects growth of excitation within a cochlear channel at or near the signal frequency.

Theories of Forward Masking

In Experiment 1, when on- and off-frequency maskers were matched in effectiveness, producing very similar signal thresholds, the addition of an identical precursor caused a divergence in signal threshold. This result is more consistent with a gain reduction theory of forward masking than additivity of masking, and supports the idea that the threshold shift with precursors used to measure GOM functions was due to cochlear gain reduction.

The delay between the onset of the precursor and signal was long enough for gain reduction to occur at the signal place (Backus and Guinan, 2006). Since gain reduction is frequency-specific for ipsilateral tone elicitors (Lilaonitkul and Guinan, 2009a), the on-frequency precursor would elicit its strongest gain reduction at or near the 4-kHz place in the cochlea, where the subjects are assumed to be listening. Because the off-frequency masker is almost an octave lower than 4 kHz, it would be processed linearly at the 4-kHz place (Ruggero et al., 1997; Cooper and Guinan, 2006). This linear processing means that the off-frequency masker has no gain to be turned down at the signal place. However, the 4-kHz masker does have gain that can be turned down at the 4-kHz place due to the presence of the precursor. This differential impact of cochlear gain reduction on the two maskers leads to reduced gain for the 4-kHz masker and no change for the 2.4-kHz masker (Kawase et al., 2000). Therefore, the 2.4-kHz masker is then more effective than the 4-kHz masker, since they were matched in effectiveness in a condition without preceding sound. This leads to higher signal thresholds for the off-frequency masker condition.

Tonal and Noise Precursor Data Support Frequency Specificity

To produce the same amount of gain reduction, the precursor level had to be higher when it was off-frequency than when it was on-frequency. This is consistent with OAE data (Lilaonitkul and Guinan, 2009b) and psychoacoustic data (Jennings and Strickland, 2010) showing frequency selectivity for the precursor. The precursor duration used in the present study is quite short, 50 ms, because this was found to be the most effective duration for an on-frequency precursor (Roverud and Strickland, 2014). For an off-frequency precursor, however, a longer duration would likely have produced more gain reduction. Roverud and Strickland (2014) modeled this as the on-frequency precursor being reduced by the gain reduction it produced, while the off-frequency precursor produced gain reduction but was not affected itself. This could be part of the reason that both OAE data and psychoacoustic data (Drga et al., 2016) show that when equal-level, long-duration elicitors are used, a broad range of elicitor frequencies are effective.

The similarity between gain reduction measured with on-frequency and BBN precursors leads to two conclusions. The first is that gain reduction masking is dominated by contributing energy in the critical bandwidth. In this way, the on-frequency tone and BBN with equal decibel level within the critical bandwidth produce similar thresholds. The second is that on-frequency tones and BBN are almost equally effective elicitors of gain reduction for the elicitor duration used here. This differs from OAE data that have suggested that noises are more robust elicitors of the ipsilateral MOCR (Lilaonitkul and Guinan, 2009a). It is possible that broadband stimuli lead to a larger change in otoacoustic emissions unrelated to greater elicitation of cochlear gain reduction. Larger bandwidth effects are also seen with contralateral elicitation (Maison et al., 2000; Lilaonitkul and Guinan, 2009a) than with ipsilateral elicitation (Lilaonitkul and Guinan, 2009a), so it may be that contralateral elicitation of cochlear gain reduction does integrate across cochlear place, but that this is not the case for ipsilateral elicitation. It is also possible that the effects depend on the signal or probe frequency; larger bandwidth effects were seen at lower frequencies using SFOAEs (Lilaonitkul and Guinan, 2009a). This could be explored in future experiments with behavioral measures.

The slope of increased gain reduction with increased BBN precursor level has implications for gain reduction research. The shallow slope means that small changes in input level of elicitor of the MOCR do not lead to large changes in cochlear gain reduction. This means that studies with differing elicitor levels can more easily be compared; differences in input level are smaller at the output of the system. However, for off-frequency elicitors, gain reduction increases with a faster rate as level is increased. This may be important in real-world situations where the noise may be low frequency.

Demonstrated Impact of Gain Reduction on Perception

This experiment demonstrated that gain reduction, measured psychophysically, grows with level and this growth varies in slope

depending on the elicitor used. The authors interpret the data as supporting that the excitation at the signal place, regardless of frequency, determines the amount of gain reduction. Different frequencies will differ in the compression applied to the input, affecting the slope of estimated gain reduction with increasing level. The similar shallow slope for increased gain reduction as a function of level with on-frequency and BBN precursors suggests that ipsilateral BBN elicitors are similarly effective to tonal elicitors, contrary to findings with SFOAE measurements (Lilaonitkul and Guinan, 2009a). The MOCR provides a mechanism for the peripheral auditory system to adaptively vary cochlear gain. This study supported that the amount of gain reduction increases with increasing level of the auditory environment, which may help the auditory system to remain sensitive to new information over the wide range of levels that we can hear.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Institutional Review Board at Purdue University. The participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

KDM was involved in research design, collected data, analyzed the data, and wrote and edited the manuscript. ES guided the research design and wrote and edited the manuscript. Both authors made significant contributions to warrant authorship and approved the final version for submission.

FUNDING

This research was supported by the National Institutes of Health (National Institute on Deafness and Other Communication Disorders, NIDCD) R01-DC008327 (ES).

REFERENCES

Backus, B. C., and Guinan, J. J. (2006). Time-course of the human medial olivocochlear reflex. *J. Acoust. Soc. Am.* 119, 2889–2904. doi: 10.1121/1.2169918

Barr, D. J., Levy, R., Scheepers, C., and Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: keep it maximal. *J. Mem. Lang.* 68, 255–278. doi: 10.1016/j.jml.2012.11.001

Bonfils, P., and Puel, J.-L. (1987). Functional properties of the crossed part of the medial olivo-cochlear bundle. *Hear. Res.* 28, 125–130. doi: 10.1016/0378-5955(87)90043-8

Bonfils, P., Remond, M.-C., and Pujol, R. (1986). Efferent tracts and cochlear frequency selectivity. *Hear. Res.* 24, 277–283. doi: 10.1016/0378-5955(86)90026-2

Cooper, N. P., and Guinan, J. J. (2003). Separate mechanical processes underlie fast and slow effects of medial olivocochlear efferent activity. *J. Physiol.* 548, 307–312. doi: 10.1111/j.1469-7793.2003.00307.x

Cooper, N. P., and Guinan, J. J. (2006). Efferent-mediated control of basilar membrane motion. *J. Physiol.* 576, 49–54. doi: 10.1113/jphysiol.2006.114991

DeRoy Milvae, K., Alexander, J. M., and Strickland, E. A. (2021). The relationship between ipsilateral cochlear gain reduction and speech-in-noise recognition at positive and negative signal-to-noise ratios. *J. Acoust. Soc. Am.* 149, 3449–3461. doi: 10.1121/10.0003964

DeRoy Milvae, K., and Strickland, E. A. (2018). Psychoacoustic measurements of ipsilateral cochlear gain reduction as a function of signal frequency. *J. Acoust. Soc. Am.* 143, 3114–3125. doi: 10.1121/1.5038254

Dolan, D. F., Guo, M. H., and Nuttall, A. L. (1997). Frequency-dependent enhancement of basilar membrane velocity during olivocochlear bundle stimulation. *J. Acoust. Soc. Am.* 102, 3587–3596. doi: 10.1121/1.421008

Drga, V., Plack, C. J., and Yasin, I. (2016). "Frequency tuning of the efferent effect on cochlear gain in humans," in *Physiology, Psychoacoustics and Cognition in Normal and Impaired Hearing*, Vol. 894, eds P. van Dijk, D. Başkent, E. Gaudrain, E. de Kleine, A. Wagner, and C. Lanting (Cham: Springer), 477–484. doi: 10.1007/978-3-319-25474-6_50

Evans, E. F. (1981). "The dynamic range problem: place and time coding at the level of cochlear nerve and nucleus," in *Neuronal Mechanisms of Hearing*, eds J. Syka and L. Aitkin (Boston, MA: Springer), 69–85. doi: 10.1007/978-1-4684-3908-3_9

Fletcher, M. D., Krumbholz, K., and de Boer, J. (2016). Effect of contralateral medial olivocochlear feedback on perceptual estimates of cochlear gain and compression. *J. Assoc. Res. Otolaryngol.* 17, 559–575. doi: 10.1007/s10162-016-0574-8

Francis, N. A., and Guinan, J. J. (2010). Acoustic stimulation of human medial olivocochlear efferents reduces stimulus-frequency and click-evoked otoacoustic emission delays: implications for cochlear filter bandwidths. *Hear. Res.* 267, 36–45. doi: 10.1016/j.heares.2010.04.009

Glasberg, B. R., and Moore, B. C. J. (1990). Derivation of auditory filter shapes from notched-noise data. *Hear. Res.* 47, 103–138. doi: 10.1016/0378-5955(90)90170-T

Guinan, J. J. (2018). Olivocochlear efferents: their action, effects, measurement and uses, and the impact of the new conception of cochlear mechanical responses. *Hear. Res.* 362, 38–47. doi: 10.1016/j.heares.2017.12.012

Guinan, J. J., Backus, B. C., Lilaonitkul, W., and Aharonson, V. (2003). Medial olivocochlear efferent reflex in humans: otoacoustic emission (OAE) measurement issues and the advantages of stimulus frequency OAEs. *J. Assoc. Res. Otolaryngol.* 4, 521–540. doi: 10.1007/s10162-002-3037-3

Guinan, J. J., and Gifford, M. L. (1988). Effects of electrical stimulation of efferent olivocochlear neurons on cat auditory-nerve fibers. II. Spontaneous rate. *Hear. Res.* 33, 115–128. doi: 10.1016/0378-5955(88)90024-X

James, A. L., Harrison, R. V., Pienkowski, M., Dajani, H. R., and Mount, R. J. (2005). Dynamics of real time DPOAE contralateral suppression in chinchillas and humans. *Int. J. Audiol.* 44, 118–129. doi: 10.1080/14992020400029996

Jennings, S. G. (2021). The role of the medial olivocochlear reflex in psychophysical masking and intensity resolution in humans: a review. *J. Neurophysiol.* 125, 2279–2308. doi: 10.1152/jn.00672.2020

Jennings, S. G., and Strickland, E. A. (2010). "The frequency selectivity of gain reduction masking: analysis using two equally-effective maskers," in *The Neurophysiological Bases of Auditory Perception*, eds E. A. Lopez-Poveda, A. R. Palmer, and R. Meddis (New York, NY: Springer), 47–58. doi: 10.1007/978-1-4419-5686-6_5

Jennings, S. G., and Strickland, E. A. (2012). Evaluating the effects of olivocochlear feedback on psychophysical measures of frequency selectivity. *J. Acoust. Soc. Am.* 132, 2483–2496. doi: 10.1121/1.4742723

Jennings, S. G., Strickland, E. A., and Heinz, M. G. (2009). Precursor effects on behavioral estimates of frequency selectivity and gain in forward masking. *J. Acoust. Soc. Am.* 125, 2172–2181. doi: 10.1121/1.3081383

Kawase, T., Ogura, M., Hidaka, H., Sasaki, N., Suzuki, Y., and Takasaka, T. (2000). Effects of contralateral noise on measurement of the psychophysical tuning curve. *Hear. Res.* 142, 63–70. doi: 10.1016/S0378-5955(00)00010-1

Krull, V., and Strickland, E. A. (2008). The effect of a precursor on growth of forward masking. *J. Acoust. Soc. Am.* 123, 4352–4357. doi: 10.1121/1.2912440

Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *J. Acoust. Soc. Am.* 49, 467–477. doi: 10.1121/1.1912375

Liberman, M. C., and Brown, M. C. (1986). Physiology and anatomy of single olivocochlear neurons in the cat. *Hear. Res.* 24, 17–36. doi: 10.1016/0378-5955(86)90003-1

Lilaonitkul, W., and Guinan, J. J. (2009a). Human medial olivocochlear reflex: effects as functions of contralateral, ipsilateral, and bilateral elicitor bandwidths. *J. Assoc. Res. Otolaryngol.* 10, 459–470. doi: 10.1007/s10162-009-0163-1

Lilaonitkul, W., and Guinan, J. J. (2009b). Reflex control of the human inner ear: a half-octave offset in medial efferent feedback that is consistent with an efferent role in the control of masking. *J. Neurophysiol.* 101, 1394–1406. doi: 10.1152/jn.90925.2008

Lilaonitkul, W., and Guinan, J. J. (2012). Frequency tuning of medial-olivocochlear-efferent acoustic reflexes in humans as functions of probe frequency. *J. Neurophysiol.* 107, 1598–1611. doi: 10.1152/jn.00549.2011

Luke, S. G. (2017). Evaluating significance in linear mixed-effects models in R. *Behav. Res. Methods* 49, 1494–1502. doi: 10.3758/s13428-016-0809-y

Maison, S., Micheyl, C., Andéol, G., Gallégo, S., and Collet, L. (2000). Activation of medial olivocochlear efferent system in humans: influence of stimulus bandwidth. *Hear. Res.* 140, 111–125. doi: 10.1016/S0378-5955(99)00196-3

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., and Bates, D. (2017). Balancing Type I error and power in linear mixed models. *J. Mem. Lang.* 94, 305–315. doi: 10.1016/j.jml.2017.01.001

Nelson, D. A., Schroder, A. C., and Wojtczak, M. (2001). A new procedure for measuring peripheral compression in normal-hearing and hearing-impaired listeners. *J. Acoust. Soc. Am.* 110, 2045–2064. doi: 10.1121/1.1404439

Norman, M., and Thornton, A. R. D. (1993). Frequency analysis of the contralateral suppression of evoked otoacoustic emissions by narrow-band noise. *Br. J. Audiol.* 27, 281–289. doi: 10.3109/03005369309076705

Oxenham, A. J., and Moore, B. C. J. (1994). Modeling the additivity of nonsimultaneous masking. *Hear. Res.* 80, 105–118. doi: 10.1016/0378-5955(94)90014-0

Oxenham, A. J., and Moore, B. C. J. (1995). Additivity of masking in normally hearing and hearing-impaired subjects. *J. Acoust. Soc. Am.* 98, 1921–1934. doi: 10.1121/1.413376

Oxenham, A. J., and Plack, C. J. (1997). A behavioral measure of basilar-membrane nonlinearity in listeners with normal and impaired hearing. *J. Acoust. Soc. Am.* 101, 3666–3675. doi: 10.1121/1.418327

Oxenham, A. J., and Plack, C. J. (2000). Effects of masker frequency and duration in forward masking: further evidence for the influence of peripheral nonlinearity. *Hear. Res.* 150, 258–266. doi: 10.1016/S0378-5955(00)00206-9

Penner, M. J., and Shiffrin, R. M. (1980). Nonlinearities in the coding of intensity within the context of a temporal summation model. *J. Acoust. Soc. Am.* 67, 617–627. doi: 10.1121/1.383885

Plack, C. J., and Oxenham, A. J. (1998). Basilar-membrane nonlinearity and the growth of forward masking. *J. Acoust. Soc. Am.* 103, 1598–1608. doi: 10.1121/1.421294

Plack, C. J., Oxenham, A. J., and Drga, V. (2006). Masking by inaudible sounds and the linearity of temporal summation. *J. Neurosci.* 26, 8767–8773. doi: 10.1523/JNEUROSCI.1134-06.2006

R Core Team (2020). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. Available online at: https://www.R-project.org/

Roverud, E., and Strickland, E. A. (2010). The time course of cochlear gain reduction measured using a more efficient psychophysical technique. *J. Acoust. Soc. Am.* 128, 1203–1214. doi: 10.1121/1.3473695

Roverud, E., and Strickland, E. A. (2014). Accounting for nonmonotonic precursor duration effects with gain reduction in the temporal window model. *J. Acoust. Soc. Am.* 135, 1321–1334. doi: 10.1121/1.4864783

Ruggero, M. A., Rich, N. C., Recio, A., Narayan, S. S., and Robles, L. (1997). Basilar-membrane responses to tones at the base of the chinchilla cochlea. *J. Acoust. Soc. Am.* 101, 2151–2163. doi: 10.1121/1.418265

Russell, I. J., and Murugasu, E. (1997). Medial efferent inhibition suppresses basilar membrane responses to near characteristic frequency tones of moderate to high intensities. *J. Acoust. Soc. Am.* 102, 1734–1738. doi: 10.1121/1.420083

Schairer, K. S., Ellison, J. C., Fitzpatrick, D., and Keefe, D. H. (2007). Wideband ipsilateral measurements of middle-ear muscle reflex thresholds in children and adults. *J. Acoust. Soc. Am.* 121, 3607–3616. doi: 10.1121/1.2722213

Strickland, E. A., Salloom, W. B., and Hegland, E. L. (2018). Evidence for gain reduction by a precursor in an on-frequency forward masking paradigm. *Acta Acust. United Acust.* 104, 809–812. doi: 10.3813/AAA.919229

Viemeister, N. F. (1988). Intensity coding and the dynamic range problem. *Hear. Res.* 34, 267–274. doi: 10.1016/0378-5955(88)90007-X

Voeten, C. C. (2020). *Buildmer: Stepwise Elimination and Term Reordering for Mixed-Effects Regression*. R Package version 1.5. Available online at: https://CRAN.R-project.org/package=buildmer (accessed September 14, 2021).

Winslow, R. L., and Sachs, M. B. (1987). Effect of electrical stimulation of the crossed olivocochlear bundle on auditory nerve response to tones in noise. *J. Neurophysiol.* 57, 1002–1021. doi: 10.1152/jn.1987.57.4.1002

Yasin, I., Drga, V., and Plack, C. J. (2014). Effect of human auditory efferent feedback on cochlear gain and compression. *J. Neurosci.* 34, 15319–15326. doi: 10.1523/JNEUROSCI.1043-14.2014

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 DeRoy Milvae and Strickland. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Neuroscience

OPINION
published: 18 October 2021
doi: 10.3389/fnins.2021.769192

Descending Control in the Auditory System: A Perspective

*Nina Kraus**

Departments of Communication Sciences, Neurobiology, Otolaryngology, Northwestern University, Evanston, IL, United States

Keywords: afferent, efferent, auditory, descending control, memory, sound, BEAMS

Once upon a time, I spent hours threading microelectrodes into auditory cortex neurons and counting spikes as the rabbits learned a tone-signaled task. Now, it is not news that cortical neurons change their tune when sound acquires relevance. But, it was still novel in the late 1970s and it was new to watch this plasticity unfold over time (Kraus and Disterhoft, 1982). It had, after all, only been a decade since the idea of a hard-wired brain was upended by observations of a remapped somatosensory cortex in monkeys (Paul et al., 1972).

In the intervening 40 years, response plasticity has been observed in nooks and crannies of the auditory pathway. Questions remain. What is the time course? How persistent are the changes? To what extent does the type of sound matter? The type of training? The nature of the task? The context? How does hearing loss alter plasticity? Or aging? How do other sensory systems contribute to auditory plasticity? What are the roles of primary and non-primary pathways? How do the afferent and efferent systems interact in effecting changes in response properties? How does response plasticity propagate throughout brain circuitry?

I recently proposed the BEAMS hypothesis (the dynamic auditory **B**rain, via **E**fferent influence, attains a new default **A**fferent state that represents **M**emory for **S**ound; Kraus, 2021a) to tackle another question: how is auditory memory reflected in neural response properties? Here, I expand on BEAMS and provide additional context to emphasize that, in my opinion, afferent processing pathways in the auditory system, via the mechanism of descending efferent control, indeed store our memory for sound.

A one-sentence theme of this special issue and of our knowledge of descending control in the auditory system could be "the hearing brain is vast." The narrowly defined hierarchy of the classical auditory pathway is firmly in the dustbin of history. How we think about sound, how we feel about sound, the movements that accompany sound, and what we experience with our other senses play key roles in shaping our auditory infrastructure (Kraus and White-Schwoch, 2015)—what I have come to call the "sound mind" (Kraus, 2021b). Our experience with sound—lifelong, in-the-moment, and all points between—leaves a legacy on this massively interconnected auditory system courtesy of the efferent system. Control of our auditory infrastructure, not only throughout the auditory system itself but by non-auditory brain systems, points to the pervasive influence of hearing to the lives of all organisms.

What might hearing without descending control look like? This is purely anecdotal, but I think it is telling. My ongoing dialog with clinicians has provided opportunities to investigate individuals with unusual medical histories. Between these case studies and our research in impaired, typical, and expert listeners, I have seen thousands of speech-evoked frequency-following responses (FFR). The single biggest, sharpest, 10-out-of-10, A-plus knockout of a response came from an individual with bilateral cortical lesions. Without cortical influence (indeed, his cortical evoked responses were entirely absent), this individual's FFR provided a glimpse of what a raw, unmodulated afferent system, left to its own devices, might look like (White-Schwoch et al., 2019).

This superb example of a lack of descending control of auditory processing serves as a cautionary tale on two fronts. First and more prosaically, it proves that bigger is not always better. But, second

OPEN ACCESS

Edited by:
David Pérez-González, University of Salamanca, Spain

Reviewed by:
Diego Elgueda, University of Chile, Chile
Victoria M. Bajo Lorenzana, University of Oxford, United Kingdom

***Correspondence:**
Nina Kraus
nkraus@northwestern.edu

Specialty section:
This article was submitted to Auditory Cognitive Neuroscience, a section of the journal Frontiers in Neuroscience

Received: *01 September 2021*
Accepted: *21 September 2021*
Published: *18 October 2021*

Citation:
Kraus N (2021) Descending Control in the Auditory System: A Perspective. Front. Neurosci. 15:769192. doi: 10.3389/fnins.2021.769192

and more importantly, it reminds us we should not take any measurement in the afferent stream at face value. The flash of color on the fMRI, the electrophysiological squiggle, the DPOAE-gram, all must be interpreted as the product of an array of influences, some far removed from the local site of recording.

The principal realm of descending control is, of course, the efferent system: how our auditory function is altered by fill-in-the-blank—our emotions, our experience, our training, our vision, our environment... I am one of the efferent system's biggest cheerleaders, having made a career of looking at how auditory processing is affected by training and experience. However, of late, I find myself becoming more and more a champion of the afferent auditory system. While the focus of this issue is the efferent system, we must not lose sight that one of its most important purposes is to influence afferent processing.

How descending influence is manifested may be surprising, as the cortical-lesion case demonstrated. But, however it manifests, the functioning of the afferent stream under efferent control, expressed in my BEAMS hypothesis, reflects our auditory memories. The lower (in terms of the peripheral-central axis) the altered response property is found, the longer-term and more persistent the memory. Our afferent auditory system retains traces of our sonic history: a shift in the default auditory processing state. That is, we can think of the afferent auditory system as a default-mode network for hearing.

One of the brain's most important jobs is prediction. Predictive coding postulates incoming information is compared to an internal template shaped by experience. If the template does not fit the incoming information, the template is adjusted accordingly (Carbajal and Malmierca, 2018; Denham and Winkler, 2020). The study of this constant dialog of efferent control and afferent processing, as exemplified in this special issue, brings us closer to solving the mystery of how these templates/memories/default modes are formed and persist—a yin and yang of *plasticity* and *stability*, eternal cordial adversaries.

There is a range of plasticity across the auditory system—a continuum between stability and mutability. And the timecourse over which the alteration of a given set-state can be achieved varies as well. Non-primary pathway neurons are inherently more flexible than primary areas (Kraus et al., 1994; Atiani et al., 2014). Relatively speaking, central areas alter their properties more readily and quickly than more peripheral areas as revealed by simultaneous recording from multiple sites during auditory learning. In ferrets, changes in non-primary auditory cortex neurons precede changes in primary auditory cortex (Elgueda et al., 2019). In humans, non-primary AC has a larger influence than primary AC on speech perception (Hamilton et al., 2021), and cortical changes precede midbrain changes as demonstrated by Skoe et al.'s contribution to this issue (Skoe et al., 2021).

Reciprocity characterizes the nature of descending and ascending pathways. Efferent control itself is mutable (for better or worse). For example, an active efferent cholinergic system confers protection against noise-induced cochlear neuropathy (Boero et al., 2018) and auditory experience seems to shield the cochlea from loud sounds (Brashears et al., 2003; Skoe and Powell, 2021). A greater understanding of these and other principles throughout the neuraxis is increasingly within our grasp as investigative techniques improve.

This special issue of *Frontiers in Neuroscience* brings us a handful of overviews of the current state of "how things work," reinforcing the general principle of reorganization of primary centers following non-primary reorganization. It also features original research that adds to our accumulated knowledge base and provides new methods to enable increased granularity as we continue to investigate the wheres and whens and hows of auditory learning in the interactive efferent and afferent streams.

The sound mind is remarkably dynamic, adaptive, and evolving. Each of us—through experience with the sounds that matter most to us—has forged a unique sound processing foundation to format our own construction of the sonic world. Our experience with sound adjusts the playing field (the afferent system), via descending control by the efferent system, so that new sound experiences are evaluated in light of our history.

AUTHOR CONTRIBUTIONS

NK conceived, wrote, and approved the submitted manuscript.

REFERENCES

Atiani, S., David, S. V., Elgueda, D., Locastro, M., Radtke-Schuller, S., Shamma, S. A., et al. (2014). Emergent selectivity for task-relevant stimuli in higher-order auditory cortex. *Neuron* 82, 486–499. doi: 10.1016/j.neuron.2014.02.029

Boero, L. E., Castagna, V. C., Di Guilmi, M. N., Goutman, J. D., Elgoyhen, A. B., and Gómez-Casati, M. E. (2018). Enhancement of the medial olivocochlear system prevents hidden hearing loss. *J. Neurosci.* 38, 7440–7451. doi: 10.1523/JNEUROSCI.0363-18.2018

Brashears, S. M., Morlet, T. G., Berlin, C. I., and Hood, L. J. (2003). Olivocochlear efferent suppression in classical musicians. *J. Am. Acad. Audiol.* 14, 314–324. doi: 10.1055/s-0040-1715747

Carbajal, G. V., and Malmierca, M. S. (2018). The neuronal basis of predictive coding long the auditory pathway: from the subcortical roots to cortical deviance detection. *Trends Hear.* 22, 1–33. doi: 10.1177/2331216518784822

Denham, S. L., and Winkler, I. (2020). Predictive coding in auditory perception: challenges and unresolved questions. *Eur. J. Neurosci.* 51, 1151–1160. doi: 10.1111/ejn.13802

Elgueda, D., Duque, D., Radtke-Schuller, S., Yin, P., David, S. V., Shamma, S. A., et al. (2019). State-dependent encoding of sound and behavioral meaning in a tertiary region of the ferret auditory cortex. *Nat. Neurosci.* 22, 447–459. doi: 10.1038/s41593-018-0317-8

Hamilton, L. S., Oganian, Y., Hall, J., and Chang, E. F. (2021). Parallel and distributed encoding of speech across human auditory cortex. *Cell* 184, 4626–4639.e13. doi: 10.1016/j.cell.2021.07.019

Kraus, N. (2021a). Memory for sound: the BEAMS hypothesis [Perspective]. *Hear. Res.* 407:108291. doi: 10.1016/j.heares.2021.108291

Kraus, N. (2021b). *Of Sound Mind: How Our Brain Constructs a Meaningful Sonic World*. Cambridge, MA: MIT Press.

Kraus, N., and Disterhoft, J. F. (1982). Response plasticity of single neurons in rabbit auditory association cortex during tone-signalled learning. *Brain Res.* 246, 205–215. doi: 10.1016/0006-8993(82)91168-4

Kraus, N., McGee, T., Littman, T., Nicol, T., and King, C. (1994). Nonprimary auditory thalamic representation of acoustic change. *J. Neurophysiol.* 72, 1270–1277. doi: 10.1152/jn.1994.72.3.1270

Kraus, N., and White-Schwoch, T. (2015). Unraveling the biology of auditory learning: a cognitive-sensorimotor-reward framework. *Trends Cogn. Sci.* 19, 642–654. doi: 10.1016/j.tics.2015.08.017

Paul, R. L., Goodman, H., and Merzenich, M. (1972). Alterations in mechanoreceptor input to Brodmann's areas 1 and 3 of the postcentral hand area of Macaca mulatta after nerve section and regeneration. *Brain Res.* 39, 1–19. doi: 10.1016/0006-8993(72)90782-2

Skoe, E., Krizman, J., Spitzer, E. R., and Kraus, N. (2021). Auditory cortical changes precede brainstem changes during rapid implicit learning: evidence from human EEG. *Front. Neurosci.* 15:718230. doi: 10.3389/fnins.2021.718230

Skoe, E., and Powell, S. (2021). *The Interaction of Musical Training and Noise Exposure on Medial Olivocochlear Efferent Function*. Association for Research in Otolaryngology MidWinter Meeting.

White-Schwoch, T., Anderson, S., Krizman, J., Nicol, T., and Kraus, N. (2019). Case studies in neuroscience: subcortical origins of the frequency-following response. *J. Neurophysiol.* 122, 844–848. doi: 10.1152/jn.00112.2019

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Kraus. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Systems Neuroscience

ORIGINAL RESEARCH
published: 21 October 2021
doi: 10.3389/fnsys.2021.674098

Patterns of Unilateral and Bilateral Projections From Layers 5 and 6 of the Auditory Cortex to the Inferior Colliculus in Mouse

Nathiya Vaithiyalingam Chandra Sekaran[1,2†], Meena S. Deshpande[1,2†], Baher A. Ibrahim[1,2], Gang Xiao[1,2], Yoshitaka Shinagawa[1,2] and Daniel A. Llano[1,2,3]*

[1] Department of Molecular and Integrative Physiology, University of Illinois at Urbana-Champaign, Champaign, IL, United States, [2] Beckman Institute for Advanced Science and Technology, Urbana, IL, United States, [3] Carle Illinois College of Medicine, Urbana, IL, United States

The auditory cortex sends massive projections to the inferior colliculus, but the organization of this pathway is not yet well understood. Previous work has shown that the corticocollicular projection emanates from both layers 5 and 6 of the auditory cortex and that neurons in these layers have different morphological and physiological properties. It is not yet known in the mouse if both layer 5 and layer 6 project bilaterally, nor is it known if the projection patterns differ based on projection location. Using targeted injections of Fluorogold into either the lateral cortex or dorsal cortex of the inferior colliculus, we quantified retrogradely labeled neurons in both the left and right lemniscal regions of the auditory cortex, as delineated using parvalbumin immunostaining. After dorsal cortex injections, we observed that approximately 18–20% of labeled cells were in layer 6 and that this proportion was similar bilaterally. After lateral cortex injections, only ipsilateral cells were observed in the auditory cortex, and they were found in both layer 5 and layer 6. The ratio of layer 5:layer 6 cells after lateral cortex injection was similar to that seen after dorsal cortex injection. Finally, injections of different tracers were made into the two inferior colliculi, and an average of 15–17% of cells in the auditory cortex were double-labeled, and these proportions were similar in layers 5 and 6. These data suggest that (1) only the dorsal cortex of the inferior colliculus receives bilateral projections from the auditory cortex, (2) both the dorsal and lateral cortex of the inferior colliculus receive similar layer 5 and layer 6 auditory cortical input, and (3) a subpopulation of individual neurons in both layers 5 and 6 branch to innervate both dorsal cortices of the inferior colliculus.

Keywords: auditory cortex, inferior colliculus, corticocollicular, corticotectal, auditory midbrain, Fluorogold, cholera toxin B, retrobeads

OPEN ACCESS

Edited by:
David K. Ryugo, Garvan Institute of Medical Research, Australia

Reviewed by:
Stephen Lomber, McGill University, Canada
Manuel S. Malmierca, University of Salamanca, Spain

***Correspondence:**
Daniel A. Llano
llano@life.illinois.edu

[†] These authors have contributed equally to this work

Received: *28 February 2021*
Accepted: *19 August 2021*
Published: *21 October 2021*

Citation:
Vaithiyalingam Chandra Sekaran N, Deshpande MS, Ibrahim BA, Xiao G, Shinagawa Y and Llano DA (2021) Patterns of Unilateral and Bilateral Projections From Layers 5 and 6 of the Auditory Cortex to the Inferior Colliculus in Mouse. Front. Syst. Neurosci. 15:674098. doi: 10.3389/fnsys.2021.674098

INTRODUCTION

The auditory corticocollicular system consists of a large set of descending projections from the auditory cortex (AC) to the inferior colliculus (IC), which is the midbrain integration center (Suga, 2008; Malmierca and Ryugo, 2011; Bajo and King, 2012; Stebbings et al., 2014). The projection primarily targets the non-lemniscal nuclei of the IC: the dorsal cortex (DC) and lateral cortex (LC)

(Saldaña et al., 1996; Winer et al., 1998; Torii et al., 2013). Stimulation and inactivation of AC inputs have been shown to have prominent effects upon the response properties of IC cells with respect to sound frequency and intensity, cues for spatial sound localization, and plastic changes in the IC (Jen et al., 1998; Yan and Suga, 1998; Popelář et al., 2003; Yan et al., 2005; Sun et al., 2007; Bajo et al., 2010; Asokan et al., 2018). The corticocollicular system is also heterogeneous and has been shown to emanate from two distinct layers of the AC: a large projection from layer 5 and a smaller (25% of the total in mice) projection from lower layer 6 (Schofield, 2009; Slater et al., 2019). Previous work has shown that neurons in these layers have different physiological properties, receive different cortical and thalamic inputs, and have different termination sizes in the IC (Slater et al., 2013, 2019; Yudintsev et al., 2019). Although the functional impact of layer 5 vs. layer 6 projections onto IC neurons is not yet known, in other corticofugal systems such as the corticothalamic projection, layer 5 and layer 6 neurons have different impacts on their synaptic targets, and likely have different roles in sensory processing (Ojima, 1994; Reichova and Sherman, 2004; Takayanagi and Ojima, 2006; Theyel et al., 2010; Williamson and Polley, 2019).

The corticocollicular system is one of multiple auditory corticofugal pathways that have cascading connectivity to ultimately influence auditory processing at the level of the auditory periphery (Xiao and Suga, 2002; Perrot et al., 2006; León et al., 2012; Terreros and Délano, 2015). Although these corticofugal systems appear to have many common organizational properties, one way in which they appear to differ is in the degree to which the projections are bilateral. Corticothalamic projections from AC to the medial geniculate body appear to be unilateral, whereas AC projections to the cochlear nucleus appear to be bilateral, but with an ipsilateral bias (Weedman and Ryugo, 1996; Jacomme et al., 2003; Schofield and Coomes, 2005). Studies have shown that corticocollicular neurons have a bilateral component in opossums, guinea pigs, and hedgehogs (Willard and Martin, 1984; Künzle, 1995; Saldaña et al., 1996; Coomes et al., 2005), with the majority of the projection being ipsilateral.

Despite the anatomical differences between layer 5 and layer 6 projections to the IC, many questions remain about the functional organization of this projection. For example, it is unknown in the mouse if both layer 5 and layer 6 project to both the DC and LC of the IC, and if so, if they project in equal proportions. In addition, it is unknown if both layer 5 and layer 6 project to both ICs in the mouse, and if so, if individual neurons branch to project to both sides. Therefore, in the current study, we injected a sensitive retrograde tracer, Fluorogold, unilaterally into either the DC or the LC of the mouse and determined the extent to which the projection was in the ipsi- vs. contralateral lemniscal AC fields. To determine if the bilaterality of the projection was due to individual neurons that branch to both ICs, or comprise separate ipsi- and contralateral projections, different tracers were placed in each IC, and examination was done for double-labeled cells in the AC. We found that while both LC and DC received layer 5 and layer 6 AC inputs in similar proportions, only the DC received bilateral inputs from the AC. In addition, we observed that after injections of different tracers to the two different ICs, double-labeled cells were observed in both layers 5 and 6, suggesting that a subset of cells from each layer branch to innervate each DC.

MATERIALS AND METHODS

Animals

Experiments were performed in adult CBA/CaJ (Jackson Labs, 000654) or Swiss Webster (Envigo, Hsd:ND4) mice of both sexes ranging from ages 4 to 6 months. Swiss Webster mice were only used for the three injections involving red retrobeads. Thirteen mice were used in the quantitative analysis of this study, and their data are summarized in **Tables 1**, **2**. Eight additional animals were excluded due to poor injection sites and were not included in the analysis. All procedures were approved by the Institutional Animal Care and Use Committee at the University of Illinois. Animals were housed in care facilities approved by the Association for Assessment and Accreditation of Laboratory Animal Care International. Every attempt was made to minimize the number of animals used and to reduce suffering at all stages of the study.

Tracer Injection

Sterile instruments and aseptic techniques were used for all surgical procedures. Mice were anesthetized intraperitoneally with a mixture of ketamine hydrochloride (100 mg/kg) and xylazine (3 mg/kg) and acepromazine (3 mg/kg). The mouse was placed into a Kopf Model 940 Small Animal Stereotaxic Instrument with digital readout. The head was shaved and disinfected with Povidone–iodide and 70% ethanol. An incision was made in the scalp and the surrounding skin was injected with lidocaine (2% Covetrus, United States) intradermally as local anesthetic and carprofen (3 mg/kg, Henry Schein Melville, NY, United States) was given subcutaneously for post-operative pain management. Moisture Eyes ophthalmic ointment was applied to each eye to protect the cornea from drying. A small craniotomy was made over the IC using a surgical drill, and a small glass micropipette (tip size approximately 10 microns) was filled with Fluorogold (Fluorochrome, LLC, Denver, CO, United States), 1% in phosphate-buffered saline (PBS) for unilateral injections, and 0.1 acetate buffer (pH 3.3) for the bilateral injections. The left side was chosen for consistency, given left–right differences that have been described in the mouse auditory system (Oviedo et al., 2010; Levy et al., 2019). For bilateral injections, the right IC was injected with either Cholera Toxin B conjugated to Alexa-Fluor 555 (CTB-555, Invitrogen Cat# C34776) at 1% in PBS, Lumafluor red retrobeads (RB, diluted 1:1.5 in PBS), or unconjugated CTB (Listlabs #104, 2 mg/mL in PBS). In all cases, tracers were pressure injected into the IC using a WPI Nanoliter 2010 injector and Micro4 pump controller at 10–20 nl per min. Volumes of injectates varied and are shown in **Tables 1**, **2**. After awakening from surgery, animals were returned to their home cages and to their vivarium until euthanized.

TABLE 1 | All animals used for unilateral injections in this study, with numbers of cells counted per animal.

			Ipsilateral				Contralateral				
Mouse	**Tracer**	**Volume injected (nL)**	**Ipsi-Layer 5**	**Ipsi-Layer 6**	**Ipsi% Layer 5**	**Ipsi% Layer 6**	**Contra-Layer 5**	**Contra-Layer 6**	**Contra% Layer 5**	**Contra% Layer 6**	**Ratio Ipsi/Contra**
M1-DC	FG	40	1971	630	75.8	24.2	330	57	85.3	14.7	6.7
M2-DC	FG	40	1611	347	82.3	17.7	162	43	79.0	21.0	9.6
M3-DC	FG	40	731	230	76.1	23.9	126	21	85.7	14.3	6.5
M4-DC	FG	40	1193	195	86.0	14.1	265	64	80.6	19.5	4.2
M1-LC	FG	40	1217	376	76.4	23.6	0	0	0	0	–
M2-LC	FG	40	701	140	84.4	16.7	0	0	0	0	–
M3-LC	FG	40	1744	458	79.2	20.8	0	0	0	0	–
M4-LC	FG	40	1149	91	92.7	7.3	0	0	0	0	–

Contra, contralateral; Ipsi, ipsilateral; FG, Fluorogold.

TABLE 2 | All animals used for bilateral injections in this study, with numbers of cells counted per animal.

Mouse	**Tracer**	**Volume injected (nL)**	**Layer 5 from Left**	**Layer 5 from Right**	**Layer 5 double label**	**Layer 5 % double labeled**	**Layer 6 from Left**	**Layer 6 from Right**	**Layer 6 double label**	**Layer 6 % double labeled**
M1-Bilat	FG Left	100	388	145	90	34.2%	93	66	29	22.3%
	CTB-555 Right	100								
M2-Bilat	FG Left	200	112	355	12	2.6%	15	17	2	6.7%
	CTB Right	300								
M4-Bilat	FG Left	200	415	285	127	22.2%	77	70	25	20.5%
	RB Right	300								
M5-Bilat	FG Left	200	523	636	111	10.6%	128	155	41	16.9%
	RB Right	300								
M6-Bilat	FG Left	200	353	574	76	8.9%	79	98	25	16.5%
	RB Right	300								

Cells were pooled from two hemispheres. FG, Fluorogold; CTB, cholera toxin B; RB, red beads.

Tissue Processing

Following a 7-day survival period, animals were anesthetized with overdose of ketamine and xylazine (200 mg/kg, 6 mg/kg) and perfused transcardially with 4% paraformaldehyde (PFA) in PBS at pH 7.4. Brains were removed and post fixed overnight in the PFA solution. After being cryoprotected in an ascending series of sucrose solutions, each brain was embedded and cut into 40–50-μm-thick coronal sections on a cryostat that were collected serially in two sets.

Immunostaining

Parvalbumin (PV) immunostaining was done to delineate the borders of the lemniscal regions of the AC [primary AC and anterior auditory field (Molinari et al., 1995; Kosaki et al., 1997; Cruikshank et al., 2001; Llano and Sherman, 2008)]. Sections were microwaved for 15 s and then incubated for 30 min in a solution of 0.3% Triton X-100 in PBS to enhance membrane permeability. The sections were then transferred to a blocking solution consisting of 0.3% Triton X-100 and 3% goat serum in PBS and incubated for 30 min. The primary antibody solution consisted of 1:500 monoclonal anti-PV raised in mouse (P 3088, Sigma Aldrich) in the blocking solution. Sections were incubated in this solution overnight and rinsed in three changes of the Triton X-100 in PBS solution the following day. The sections were then transferred to a secondary antibody solution and incubated at room temperature for 2 h. This solution consisted of 1:100 Alexa Fluor 568-conjugated goat anti-mouse secondary antibody (catalog #A-11004, Invitrogen). Following a final series of washes in PBS, the sections were mounted on gelatin-coated slides and coverslipped with an anti-fade solution (Vectashield; Vector Laboratories). For immunostaining of CTB, after blocking with 3% donkey serum in 0.3% Triton X-100, sections were incubated with 1:10,000 anti-CTB antibody in blocking solution (#703 Listlabs) overnight at 4°C. After washing, sections were incubated with 1:200 solution of Donkey anti-Goat IgG (H + L) Cross-Adsorbed Secondary Antibody conjugated to Alexa Fluor 555 (Invitrogen # A-21432) diluted in blocking buffer with a 1:200 dilution.

Imaging and Analysis

Sections were imaged with a Leica SP8 laser scanning confocal microscope and LAS X control software or with an Olympus IX71 inverted epifluorescence microscope. For confocal images, each IC tissue section containing retrograde label, 40 × mosaic Z-stacks were taken throughout the entire depth and *x–y* plane of the IC. The stacks were collapsed into 2D maximum intensity projections and tiled into a single image using LAS X software. ImageJ software was used to adjust the color balance and

to draw masks around the edge of the tissue to remove the embedding medium.

Data Quantification and Statistics

PV immunostaining was used to mark the borders of the lemniscal portions of the AC. For counting Fluorogold-labeled cells in the AC, neurons from six non-consecutive sections were selected. These sections were distributed across the anterior–posterior dimension of the AC. The labeled cells in AC within the PV-enriched zone were counted using ImageJ software. For counting double-labeled cells, to ensure that areas of maximum overlap were examined, only sections where at least 20 cells per tracer type were present were analyzed. Given the small numbers of animals ($n = 4$ in each injection location for unilateral injections, $n = 5$ for bilateral tracer injections), normality was not assumed and non-parametric statistics were used throughout with a threshold for significance of $p < 0.05$. Data are presented as median ± standard deviation.

RESULTS

Thirteen adult (4–6 months old) male and female mice were used in this study. Four were injected with Fluorogold into the left LC, four were injected into the left DC, and five were injected into both ICs. For both DC and LC injection sites, there tended to be some spillover into the central nucleus of the IC (CNIC). See **Tables 1**, **2** for a listing of all mice used for quantitative analysis in this study. For the mice injected into either DC or LC, PV immunostaining was done to delineate the lemniscal regions of the AC. An example of a DC injection site as well as images from both ACs demonstrating layer 5 and layer 6 retrograde label, and the corresponding PV images are shown in **Figure 1**.

Distributions of Ipsi- vs. Contralateral Cells in Auditory Cortex After Dorsal Cortex or Lateral Cortex Tracer Injection

The total number of cells in the lemniscal regions of the ipsilateral and contralateral AC were counted and compared. In all cases of DC injections ($n = 4$), there were significantly greater numbers of retrogradely labeled cells in the ipsilateral cortex ($p = 0.02$, Wilcoxon signed-rank test), with a median ipsilateral:contralateral ratio = 6.63 ± 2.18, indicating that approximately 85% of all labeled cells were ipsilateral. After injection into the LC ($n = 4$), 100% of all labeled cells in all animals were found in the ipsilateral AC.

Distribution of Layer 5 vs. Layer 6 Corticocollicular Cells in Ipsi- vs. Contralateral Dorsal Cortex Injection

After DC injection, the proportions of layer 5 vs. layer 6 corticocollicular cells in the lemniscal AC were compared using a within-animal comparison. Although the total number of cells was significantly greater on the ipsilateral side, there was no difference in the proportion of layer 6 cells on the two sides (ipsilateral: 20.83 ± 4.96% vs. contralateral: 17.09 ± 3.36%, $p = 0.69$, Wilcoxon signed-rank test, **Figure 2**). Although LC injections did not produce any contralateral label, the retrogradely labeled cells in the ipsilateral lemniscal AC contained substantial layer 6 label. We compared the proportion of ipsilateral layer 6 label after DC compared to LC injections and found no significant difference between the two (DC 20.83 ± 4.96%, LC 18.72 ± 7.11%, $p = 0.49$, Mann–Whitney).

Double-Retrograde Injection Into Left and Right Lateral Cortex

CTB-555, unconjugated CTB, or RB were injected into the right IC and Fluorogold was injected into the left IC, and images containing cells stained with each tracer were overlaid. Similar to previous findings (Coomes et al., 2005), we found three cell labeling types in each layer: CTB- or RB-only labeled cells, Fluorogold-only labeled cells, and cells labeling with both a red tracer and Fluorogold. All three cell types were found in both layer 5 and layer 6 (**Figure 3**), suggesting the presence of branching cells in each layer. The numbers of single- and double-labeled cells in each layer were counted and summarized in **Table 2**. The proportions of double-labeled cells ranged from

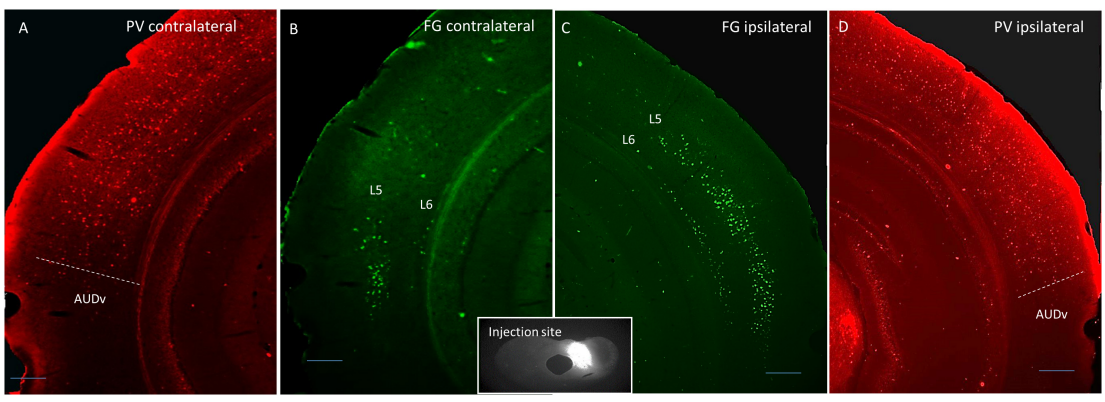

FIGURE 1 | Coronal images showing an example DC injection site (inset), and the corresponding contralateral Fluorogold label **(B)**, ipsilateral Fluorogold label **(C)**, and PV immunostaining pattern **(A,D)**. Scale bar = 250 μm. L5 = layer 5, L6 = layer 6, AUDv, ventral auditory region.

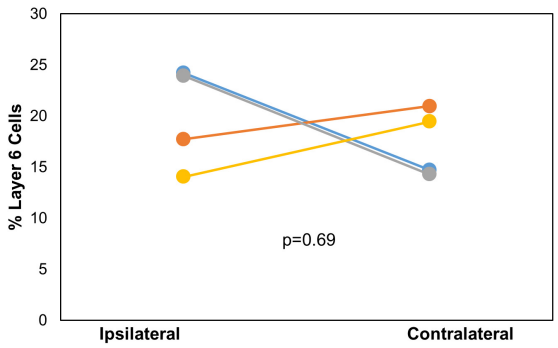

FIGURE 2 | Comparison of the layer 6 percent of total cells located ipsilaterally and contralaterally of the lemniscal AC between each mouse (n = 4) with injection site in DC. p-value calculated using Wilcoxon Signed Rank Test. See **Table 1** for actual values. n = 4 mice. Each color corresponds to a different mouse.

2.6 to 34.2% per layer per animal. The mean proportion of double-labeled cells was similar in layer 5 compared to layer 6 (15.7 ± 12.5 [SD]% in layer 5 vs. 16.6 ± 6.1 [SD]% in layer 6, n = 5 mice, p = 0.686, Wilcoxon Signed Rank, see **Figure 4**).

DISCUSSION

In this report, we used single- and double-retrograde tracing methods in the mouse to investigate patterns of unilateral vs. bilateral input from AC layers 5 and 6 to the IC. We report three main findings: (1) only the DC receives bilateral projections from the AC, (2) both the DC and LC receive similar proportions of layer 5 and layer 6 auditory cortical input, and (3) a subset of individual neurons in both layers 5 and 6 branch to innervate the DC bilaterally. These findings are summarized in the model shown in **Figure 5**. Below, we discuss the implications of these findings.

Technical Considerations

Cortical areas in this study were defined using PV immunostaining, which has been established to distinguish lemniscal auditory areas (primary AC and anterior auditory field) from non-lemniscal areas (Molinari et al., 1995; Kosaki et al., 1997; Cruikshank et al., 2001). Thus, we have not attempted to differentiate primary AC vs. anterior auditory field proportions of either layer 5 or layer 6, although we consistently see similar proportions of each cell type throughout the anterior–posterior extent of the PV-enriched zone. Future work using *in vivo* mapping of auditory fields prior to the injection of tracers may be helpful to determine if differences exist in the layer 5 vs. 6 projections from primary AC or anterior auditory field to the IC, or in the non-lemniscal regions of the auditory cortex.

The proportion of cells found to be double-labeled after injection of different tracers into different sides was found to be relatively low (range = 2.6–34.2%). Previous work has shown that this approach is susceptible to significant undercounting (Doucet et al., 2003; Coomes et al., 2005; Schofield et al., 2007). For example, when co-injecting mixed tracers of different chemical entities, one of which being synthetic beads (similar to the current study) into the same location of the IC, a range of 4–70.1% of AC cells were double-labeled (Schofield et al., 2007). The undercounting occurs because the two injections may not be matched in terms of their projection fields and because of differential efficiency of the two labels. We attempted to make our sites large enough to encompass major portions of the DC bilaterally, but without being so large as to risk entering neighboring structures, such as superior colliculus. Thus, we assume that the percentage of cells that branch to innervate both colliculi is higher than the proportion of double-labeled cells reported here.

Implications

The results of this study suggest that both the DC and LC receive similar proportions of layer 5 and layer 6 input, but that only DC receives input from the contralateral AC. These results differ somewhat from those of Schofield (2009), who did not observe contralateral corticocollicular projections from layer 6 in guinea pig. The differences may be related to the different species or tracer used. Although layer 6 projections to the IC have been identified in multiple species including mice, rats, gerbils, ferrets, and hedgehog tenrec (Games and Winer, 1988;

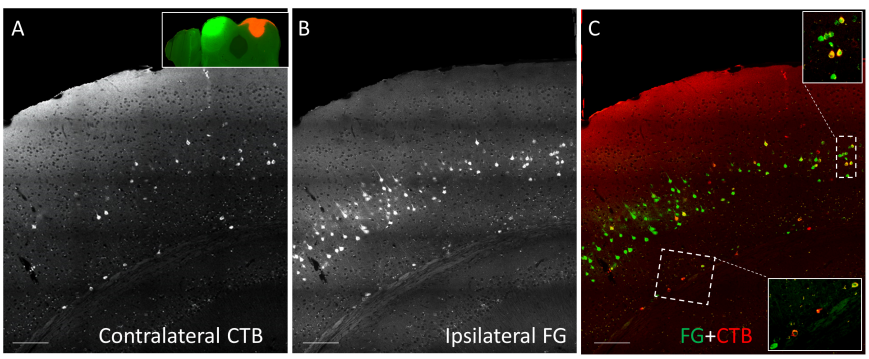

FIGURE 3 | Coronal sections showing double injection site of Fluorogold (FG) on the left and CTB-555 on the right IC. **(A)** Grayscale image showing contralateral CTB-555-labeled cells. **(B)** Grayscale image showing ipsilateral FG-labeled cells. **(C)** Overlay between panel **(A)** and panel **(B)**, showing single-labeled (red or green) and double-labeled (yellow) cells in each layer in insets. Scale bar = 100 μm.

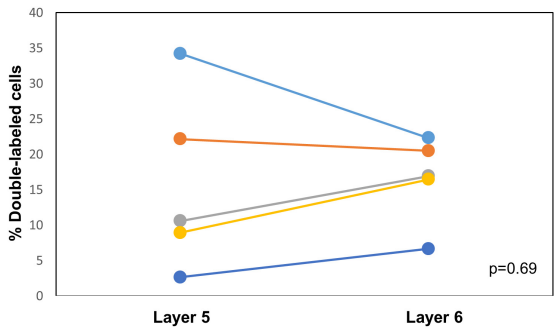

FIGURE 4 | Comparison of the proportion of double-labeled cells found in each layer after injection of different tracers into the left and right IC. See **Table 2** for actual values. $n = 5$ mice. p-value calculated using Wilcoxon Signed Rank Test. Each color corresponds to a different mouse.

Künzle, 1995; Doucet et al., 2003; Bajo and Moore, 2005; Bajo et al., 2007), the relative proportions of those projections may differ. In the case of guinea pigs, approximately 10.2% of the total cell population was determined to come from layer 6 using a variety of tracers excluding Fluorogold. However, the current study and a previous study have established that this number is roughly 20–25% in mice using Fluorogold, which is a very sensitive retrograde tracer (Schofield, 2008). This difference in the tracers, and the approximately sixfold greater ipsilateral- vs. contralateral-projecting corticocollicular cells than contralateral-projecting cells, coupled with a smaller proportion of layer 6 cells in guinea pigs, suggest that the observation of the lack of layer 6 contralateral-projecting cells in guinea pig was due to a threshold effect.

We also observed that a small proportion of cells branched to innervate both ICs (15.7% in layer 5 and 16.6% in layer 6). These values are higher than those seen in the guinea pig layer 5 [range = 2.5–11.9% (Coomes et al., 2005)]. However, given the uncertainties regarding the precise values of double-labeled cells as outlined by several authors previously (Doucet et al., 2003; Coomes et al., 2005; Schofield et al., 2007) as well as the species and tracer differences in this study, it is not clear that the differences between our study and the Coomes et al. (2005) study represent real biological differences in branching patterns. However, both studies do indicate that bilateral coordination of IC modulation is an important feature in a subset of the corticocollicular projection, and the current report extends this finding to layer 6.

The implications of having bilateral projections from AC to portions of the IC are not yet known. The IC receives bilateral projections from the auditory brainstem (Coleman and Clerici, 1987; Cant and Benson, 2006, 2008) and receives inputs from the superior olive (Kelly et al., 1998; Loftus et al., 2004), which itself gets bilateral input. Therefore, it is unlikely that a bilateral descending projection is required to produce sensitivity to sounds from both ears. The layer 5 corticofugal system has been hypothesized to serve as a system that drives rapid escape behaviors (Xiong et al., 2015; Li et al., 2021). In addition, the layer 5 corticofugal projections in other sensory systems appear to be widely branching to multiple sensory and motor regions (Deschênes et al., 1994; Bourassa and Deschenes, 1995; Bourassa et al., 1995; Kita and Kita, 2012;

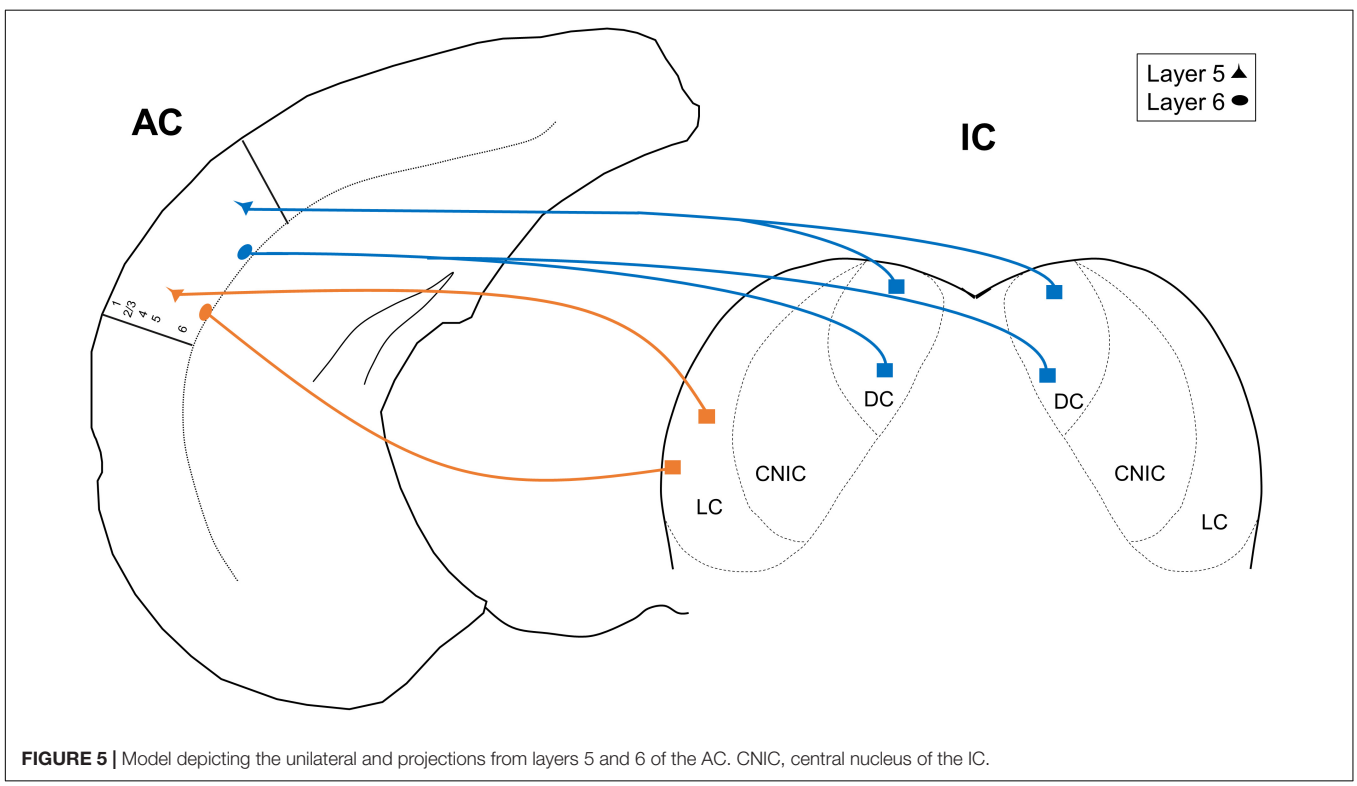

FIGURE 5 | Model depicting the unilateral and projections from layers 5 and 6 of the AC. CNIC, central nucleus of the IC.

Guo et al., 2017; Prasad et al., 2020). Therefore, it is not surprising to find a bilateral projection system to the inferior colliculi, which also project to motor structures to mediate escape responses (Kawamura, 1975; Aitkin and Boyd, 1978; Edwards et al., 1979; Cadusseau and Roger, 1985; Appell and Behan, 1990; Huffman and Henson, 1990; Lesicko and Llano, 2020). The presence of a bilateral layer 6 system, which we have previously speculated to serve a modulatory role (Yudintsev et al., 2019; Asilador and Llano, 2020), was less expected, but may suggest that the dual functions of layer 5 and layer 6 are necessary for the potential escape function of the layer 5 corticofugal projections. The absence, then, of a bilateral projection to the LC may suggest that this region is less likely to be involved in rapid motor escape behaviors than DC. Future work comparing the corticofugal response properties of cortical-recipient cells in LC or DC will help to clarify their separate roles in acoustic behavior.

Summary and Conclusion

In this study, we observed that lemniscal regions of the AC send bilateral projections from layers 5 and 6 to the DC of both inferior colliculi, with the majority being ipsilateral. We also observed that the LC receives only an ipsilateral projection from the AC and that this projection is derived from both layers 5 and 6. The proportion of layer 6 cells projecting to the IC is approximately 18–20% and does not differ based on IC target. Finally, we observed that the bilateral projection to the DC comprises, at least in part, individual neurons in both layers 5 and 6 that branch to innervate both DCs. Understanding the implications of these findings requires further investigation but may relate to the suspected roles of corticofugal projections in rapid acoustic escape behaviors.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

ETHICS STATEMENT

The animal study was reviewed and approved by Illinois Institutional Animal Care and Use Committee (IACUC).

AUTHOR CONTRIBUTIONS

NV, MD, and DL analyzed the data and co-wrote the manuscript. BI performed stereotactic injections. GX and YS processed the tissue sections. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by R01DC013073 and R01DC016599 to DL and an award from the Benjamin R. and Elinor W. Bullock and Edwin E. and Jeanne Bullock Professorial Scholar fund to DL.

REFERENCES

Aitkin, L., and Boyd, J. (1978). Acoustic input to the lateral pontine nuclei. *Hear. Res.* 1, 67–77. doi: 10.1016/0378-5955(78)90010-2

Appell, P. P., and Behan, M. (1990). Sources of subcortical GABAergic projections to the superior colliculus in the cat. *J. Comp. Neurol.* 302, 143–158. doi: 10.1002/cne.903020111

Asilador, A., and Llano, D. A. (2020). Top-down inference in the auditory system: potential roles for corticofugal projections. *Front. Neural Circuits* 14:615259.

Asokan, M. M., Williamson, R. S., Hancock, K. E., and Polley, D. B. (2018). Sensory overamplification in layer 5 auditory corticofugal projection neurons following cochlear nerve synaptic damage. *Nat. Commun.* 9:2468.

Bajo, V. M., and King, A. J. (2012). Cortical modulation of auditory processing in the midbrain. *Front. Neural Circuits* 6:114.

Bajo, V. M., and Moore, D. R. (2005). Descending projections from the auditory cortex to the inferior colliculus in the gerbil, *Meriones unguiculatus*. *J. Comp. Neurol.* 486, 101–116. doi: 10.1002/cne.20542

Bajo, V. M., Nodal, F. R., Bizley, J. K., Moore, D. R., and King, A. J. (2007). The ferret auditory cortex: descending projections to the inferior colliculus. *Cereb. Cortex* 17, 475–491. doi: 10.1093/cercor/bhj164

Bajo, V. M., Nodal, F. R., Moore, D. R., and King, A. J. (2010). The descending corticocollicular pathway mediates learning-induced auditory plasticity. *Nat. Neurosci.* 13, 253–260. doi: 10.1038/nn.2466

Bourassa, J., and Deschenes, M. (1995). Corticothalamic projections from the primary visual cortex in rats: a single fiber study using biocytin as an anterograde tracer. *Neuroscience* 66, 253–263. doi: 10.1016/0306-4522(95)00009-8

Bourassa, J., Pinault, D., and Deschenes, M. (1995). Corticothalamic projections from the cortical barrel field to the somatosensory thalamus in rats: a single-fibre study using biocytin as an anterograde tracer. *Eur. J. Neurosci.* 7, 19–30. doi: 10.1111/j.1460-9568.1995.tb01016.x

Cadusseau, J., and Roger, M. (1985). Afferent projections to the superior colliculus in the rat, with special attention to the deep layers. *J. fur Hirnforschung* 26, 667–681.

Cant, N. B., and Benson, C. G. (2006). Organization of the inferior colliculus of the gerbil (*Meriones unguiculatus*): differences in distribution of projections from the cochlear nuclei and the superior olivary complex. *J. Comp. Neurol.* 495, 511–528. doi: 10.1002/cne.20888

Cant, N. B., and Benson, C. G. (2008). Organization of the inferior colliculus of the gerbil (*Meriones unguiculatus*): projections from the cochlear nucleus. *Neuroscience* 154, 206–217. doi: 10.1016/j.neuroscience.2008.02.015

Coleman, J. R., and Clerici, W. J. (1987). Sources of projections to subdivisions of the inferior colliculus in the rat. *J. Comp. Neurol.* 262, 215–226. doi: 10.1002/cne.902620204

Coomes, D. L., Schofield, R. M., and Schofield, B. R. (2005). Unilateral and bilateral projections from cortical cells to the inferior colliculus in guinea pigs. *Brain Res.* 1042, 62–72. doi: 10.1016/j.brainres.2005.02.015

Cruikshank, S. J., Killackey, H. P., and Metherate, R. (2001). Parvalbumin and calbindin are differentially distributed within primary and secondary subregions of the mouse auditory forebrain. *Neuroscience* 105, 553–569. doi: 10.1016/s0306-4522(01)00226-3

Deschênes, M., Bourassa, J., and Pinault, D. (1994). Corticothalamic projections from layer V cells in rat are collaterals of long-range corticofugal axons. *Brain Res.* 664, 215–219. doi: 10.1016/0006-8993(94)91974-7

Doucet, J., Molavi, D., and Ryugo, D. (2003). The source of corticocollicular and corticobulbar projections in area Te1 of the rat. *Exp. Brain Res.* 153, 461–466. doi: 10.1007/s00221-003-1604-4

Edwards, S. B., Ginsburgh, C. L., Henkel, C. K., and Stein, B. E. (1979). Sources of subcortical projections to the superior colliculus in the cat. *J. Comp. Neurol.* 184, 309–329. doi: 10.1002/cne.901840207

Games, K. D., and Winer, J. A. (1988). Layer V in rat auditory cortex: projections to the inferior colliculus and contralateral cortex. *Hear. Res.* 34, 1–25. doi: 10.1016/0378-5955(88)90047-0

Guo, C., Peng, J., Zhang, Y., Li, A., Li, Y., Yuan, J., et al. (2017). Single-axon level morphological analysis of corticofugal projection neurons in mouse barrel field. *Sci. Rep.* 7:2846.

Huffman, R. F., and Henson, O. (1990). The descending auditory pathway and acousticomotor systems: connections with the inferior colliculus. *Brain Res. Rev.* 15, 295–323. doi: 10.1016/0165-0173(90)90005-9

Jacomme, A.-V., Nodal, F., Bajo, V., Manunta, Y., Edeline, J.-M., Babalian, A., et al. (2003). The projection from auditory cortex to cochlear nucleus in guinea pigs: an in vivo anatomical and in vitro electrophysiological study. *Exp. Brain Res.* 153, 467–476. doi: 10.1007/s00221-003-1606-2

Jen, P.-S., Chen, Q., and Sun, X. (1998). Corticofugal regulation of auditory sensitivity in the bat inferior colliculus. *J. Comp. Physiol. A* 183, 683–697. doi: 10.1007/s003590050291

Kawamura, K. (1975). The pontine projection from the inferior colliculus in the cat. an experimental anatomical study. *Brain Res.* 95, 309–322. doi: 10.1016/0006-8993(75)90109-2

Kelly, J. B., Liscum, A., Van Adel, B., and Ito, M. (1998). Projections from the superior olive and lateral lemniscus to tonotopic regions of the rat's inferior colliculus. *Hear. Res.* 116, 43–54.

Kita, T., and Kita, H. (2012). The subthalamic nucleus is one of multiple innervation sites for long-range corticofugal axons: a single-axon tracing study in the rat. *J. Neurosci.* 32, 5990–5999. doi: 10.1523/jneurosci.5717-11.2012

Kosaki, H., Hashikawa, T., He, J., and Jones, E. (1997). Tonotopic organization of auditory cortical fields delineated by parvalbumin immunoreactivity in macaque monkeys. *J. Comp. Neurol.* 386, 304–316. doi: 10.1002/(sici)1096-9861(19970922)386:2<304::aid-cne10>3.0.co;2-k

Künzle, H. (1995). Regional and laminar distribution of cortical neurons projecting to either superior or inferior colliculus in the hedgehog tenrec. *Cereb. Cortex* 5, 338–352. doi: 10.1093/cercor/5.4.338

León, A., Elgueda, D., Silva, M. A., Hamamé, C. M., and Délano, P. H. (2012). Auditory cortex basal activity modulates cochlear responses in chinchillas. *PLoS One* 7:e36203. doi: 10.1371/journal.pone.0036203

Lesicko, A. M., and Llano, D. A. (2020). Circuit mechanisms underlying the segregation and integration of parallel processing streams in the inferior colliculus. *J. Neurosci.* 40, 6328–6344. doi: 10.1523/jneurosci.0646-20.2020

Levy, R. B., Marquarding, T., Reid, A. P., Pun, C. M., Renier, N., and Oviedo, H. V. (2019). Circuit asymmetries underlie functional lateralization in the mouse auditory cortex. *Nat. Commun.* 10:2783.

Li, Z., Wei, J.-X., Zhang, G.-W., Huang, J. J., Zingg, B., Wang, X., et al. (2021). Corticostriatal control of defense behavior in mice induced by auditory looming cues. *Nat. Commun.* 12:1040.

Llano, D. A., and Sherman, S. M. (2008). Evidence for nonreciprocal organization of the mouse auditory thalamocortical-corticothalamic projection systems. *J. Comp. Neurol.* 507, 1209–1227. doi: 10.1002/cne.21602

Loftus, W. C., Bishop, D. C., Saint Marie, R. L., and Oliver, D. L. (2004). Organization of binaural excitatory and inhibitory inputs to the inferior colliculus from the superior olive. *J. Comp. Neurol.* 472, 330–344. doi: 10.1002/cne.20070

Malmierca, M. S., and Ryugo, D. K. (2011). "Descending connections of auditory cortex to the midbrain and brain stem," in *The Auditory Cortex*, eds J. A. Winer and C. E. Schreiner (Berlin: Springer), 189–208. doi: 10.1007/978-1-4419-0074-6_9

Molinari, M., Dell'anna, M., Rausell, E., Leggio, M., Hashikawa, T., and Jones, E. (1995). Auditory thalamocortical pathways defined in monkeys by calcium-binding protein immunoreactivity. *J. Comp. Neurol.* 362, 171–194. doi: 10.1002/cne.903620203

Ojima, H. (1994). Terminal morphology and distribution of corticothalamic fibers originating from layers 5 and 6 of cat primary auditory cortex. *Cereb. Cortex* 4, 646–663. doi: 10.1093/cercor/4.6.646

Oviedo, H. V., Bureau, I., Svoboda, K., and Zador, A. M. (2010). The functional asymmetry of auditory cortex is reflected in the organization of local cortical circuits. *Nat. Neurosci.* 13, 1413–1420. doi: 10.1038/nn.2659

Perrot, X., Ryvlin, P., Isnard, J., Guènot, M., Catenoix, H., Fischer, C., et al. (2006). Evidence for corticofugal modulation of peripheral auditory activity in humans. *Cereb. Cortex* 16, 941–948. doi: 10.1093/cercor/bhj035

Popelář, J., Nwabueze-Ogbo, F., and Syka, J. (2003). Changes in neuronal activity of the inferior colliculus in rat after temporal inactivation of the auditory cortex. *Physiol. Res.* 52, 615–628.

Prasad, J. A., Carroll, B. J., and Sherman, S. M. (2020). Layer 5 corticofugal projections from diverse cortical areas: variations on a pattern of thalamic and extrathalamic targets. *J. Neurosci.* 40, 5785–5796. doi: 10.1523/jneurosci.0529-20.2020

Reichova, I., and Sherman, S. M. (2004). Somatosensory corticothalamic projections: distinguishing drivers from modulators. *J. Neurophysiol.* 92, 2185–2197. doi: 10.1152/jn.00322.2004

Saldaña, E., Feliciano, M., and Mugnaini, E. (1996). Distribution of descending projections from primary auditory neocortex to inferior colliculus mimics the topography of intracollicular projections. *J. Comp. Neurol.* 371, 15–40. doi: 10.1002/(sici)1096-9861(19960715)371:1<15::aid-cne2>3.0.co;2-o

Schofield, B. R. (2008). Retrograde axonal tracing with fluorescent markers. *Curr. Protoc. Neurosci.* Chapter 1:Unit 1.17. doi: 10.1002/0471142301.ns0117s43

Schofield, B. R. (2009). Projections to the inferior colliculus from layer VI cells of auditory cortex. *Neuroscience* 159, 246–258. doi: 10.1016/j.neuroscience.2008.11.013

Schofield, B. R., and Coomes, D. L. (2005). Auditory cortical projections to the cochlear nucleus in guinea pigs. *Hear. Res.* 199, 89–102. doi: 10.1016/j.heares.2004.08.003

Schofield, B. R., Schofield, R. M., Sorensen, K. A., and Motts, S. D. (2007). On the use of retrograde tracers for identification of axon collaterals with multiple fluorescent retrograde tracers. *Neuroscience* 146, 773–783. doi: 10.1016/j.neuroscience.2007.02.026

Slater, B. J., Sons, S. K., Yudintsev, G., Lee, C. M., and Llano, D. A. (2019). Thalamocortical and intracortical inputs differentiate layer-specific mouse auditory corticocollicular neurons. *J. Neurosci.* 39, 256–270. doi: 10.1523/jneurosci.3352-17.2018

Slater, B. J., Willis, A. M., and Llano, D. A. (2013). Evidence for layer-specific differences in auditory corticocollicular neurons. *Neuroscience* 229, 144–154. doi: 10.1016/j.neuroscience.2012.10.053

Stebbings, K. A., Lesicko, A. M., and Llano, D. A. (2014). The auditory corticocollicular system: Molecular and circuit-level considerations. *Hear. Res.* 314, 51–59. doi: 10.1016/j.heares.2014.05.004

Suga, N. (2008). Role of corticofugal feedback in hearing. *J. Comp. Physiol. A* 194, 169–183. doi: 10.1007/s00359-007-0274-2

Sun, X., Xia, Q., Lai, C. H., Shum, D. K. Y., Chan, Y. S., and He, J. (2007). Corticofugal modulation of acoustically induced Fos expression in the rat auditory pathway. *J. Comp. Neurol.* 501, 509–525. doi: 10.1002/cne.21249

Takayanagi, M., and Ojima, H. (2006). Microtopography of the dual corticothalamic projections originating from domains along the frequency axis of the cat primary auditory cortex. *Neuroscience* 142, 769–780. doi: 10.1016/j.neuroscience.2006.06.048

Terreros, G., and Délano, P. H. (2015). Corticofugal modulation of peripheral auditory responses. *Front. Syst. Neurosci.* 9:134.

Theyel, B. B., Llano, D. A., and Sherman, S. M. (2010). The corticothalamocortical circuit drives higher-order cortex in the mouse. *Nat. Neurosci.* 13, 84–88. doi: 10.1038/nn.2449

Torii, M., Hackett, T. A., Rakic, P., Levitt, P., and Polley, D. B. (2013). EphA signaling impacts development of topographic connectivity in auditory corticofugal systems. *Cereb. Cortex* 23, 775–785. doi: 10.1093/cercor/hs066

Weedman, D. L., and Ryugo, D. K. (1996). Projections from auditory cortex to the cochlear nucleus in rats: synapses on granule cell dendrites. *J. Comp. Neurol.* 371, 311–324. doi: 10.1002/(sici)1096-9861(19960722)371:2<311::aid-cne10>3.0.co;2-v

Willard, F., and Martin, G. (1984). Collateral innervation of the inferior colliculus in the North American opossum: a study using fluorescent markers in a double-labeling paradigm. *Brain Res.* 303, 171–182. doi: 10.1016/0006-8993(84)90225-7

Williamson, R. S., and Polley, D. B. (2019). Parallel pathways for sound processing and functional connectivity among layer 5 and 6 auditory corticofugal neurons. *Elife* 8:e42974.

Winer, J. A., Larue, D. T., Diehl, J. J., and Hefti, B. J. (1998). Auditory cortical projections to the cat inferior colliculus. *J. Comp. Neurol.* 400, 147–174. doi: 10.1002/(sici)1096-9861(19981019)400:2<147::aid-cne1>3.0.co;2-9

Xiao, Z., and Suga, N. (2002). Modulation of cochlear hair cells by the auditory cortex in the mustached bat. *Nat. Neurosci.* 5, 57–63. doi: 10.1038/nn786

Xiong, X. R., Liang, F., Zingg, B., Ji, X.-Y., Ibrahim, L. A., Tao, H. W., et al. (2015). Auditory cortex controls sound-driven innate defense behaviour through corticofugal projections to inferior colliculus. *Nat. Commun.* 6:7224.

Yan, J., Zhang, Y., and Ehret, G. (2005). Corticofugal shaping of frequency tuning curves in the central nucleus of the inferior colliculus of mice. *J. Neurophysiol.* 93, 71–83. doi: 10.1152/jn.00348.2004

Yan, W., and Suga, N. (1998). Corticofugal modulation of the midbrain frequency map in the bat auditory system. *Nat. Neurosci.* 1, 54–58. doi: 10.103 8/255

Yudintsev, G., Asilador, A., Coppinger, M., Nair, K., Prasad, M., and Llano, D. A. (2019). Connectional heterogeneity in the mouse auditory corticocollicular system. *BioRxiv [Preprint]* doi: 10.1101/571711

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Vaithiyalingam Chandra Sekaran, Deshpande, Ibrahim, Xiao, Shinagawa and Llano. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 28 October 2021
doi: 10.3389/fnins.2021.746821

A Time-Course-Based Estimation of the Human Medial Olivocochlear Reflex Function Using Clicks

Sriram Boothalingam[1,2]*, *Shawn S. Goodman*[3], *Hilary MacCrae*[4] and *Sumitrajit Dhar*[4,5]

[1] Department of Communication Sciences and Disorders, University of Wisconsin-Madison, Madison, WI, United States, [2] Waisman Center, University of Wisconsin-Madison, Madison, WI, United States, [3] Department of Communication Sciences and Disorders, University of Iowa, Iowa City, IA, United States, [4] Roxelyn and Richard Pepper Department of Communication Sciences and Disorders, Northwestern University, Evanston, IL, United States, [5] Knowles Center, Northwestern University, Evanston, IL, United States

The auditory efferent system, especially the medial olivocochlear reflex (MOCR), is implicated in both typical auditory processing and in auditory disorders in animal models. Despite the significant strides in both basic and translational research on the MOCR, its clinical applicability remains under-utilized in humans due to the lack of a recommended clinical method. Conventional tests employ broadband noise in one ear while monitoring change in otoacoustic emissions (OAEs) in the other ear to index efferent activity. These methods, (1) can only assay the contralateral MOCR pathway and (2) are unable to extract the kinetics of the reflexes. We have developed a method that re-purposes the same OAE-evoking click-train to also concurrently elicit bilateral MOCR activity. Data from click-train presentations at 80 dB peSPL at 62.5 Hz in 13 young normal-hearing adults demonstrate the feasibility of our method. Mean MOCR magnitude (1.7 dB) and activation time-constant (0.2 s) are consistent with prior MOCR reports. The data also suggest several advantages of this method including, (1) the ability to monitor MEMR, (2) obtain both magnitude and kinetics (time constants) of the MOCR, (3) visual and statistical confirmation of MOCR activation.

Keywords: medial olivocochlear reflex, middle ear muscle reflex, click-evoked otoacoustic emissions, time-course, kinetics

OPEN ACCESS

Edited by:
Enrique A. Lopez-Poveda, University of Salamanca, Spain

Reviewed by:
Skyler G. Jennings, The University of Utah, United States
Miriam I. Marrufo Pérez, University of Salamanca, Spain

***Correspondence:**
Sriram Boothalingam
boothalingam@wisc.edu

Specialty section:
This article was submitted to Auditory Cognitive Neuroscience, a section of the journal Frontiers in Neuroscience

Received: *24 July 2021*
Accepted: *28 September 2021*
Published: *28 October 2021*

Citation:
Boothalingam S, Goodman SS, MacCrae H and Dhar S (2021) A Time-Course-Based Estimation of the Human Medial Olivocochlear Reflex Function Using Clicks. Front. Neurosci. 15:746821. doi: 10.3389/fnins.2021.746821

1. INTRODUCTION

The auditory efferent system serves as a dynamic feedback mechanism through which the brain regulates afferent neural inputs. Such feedback control occurs at multiple stages in the auditory system and is thought to aid in automatic and attention-driven signal detection in noise (Winslow and Sachs, 1988; de Boer and Thornton, 2007; Delano et al., 2007; Mertes et al., 2019) and protection of peripheral sensory cells from acoustic overexposure (Galambos and Rupert, 1959; Borg et al., 1983; Liberman, 1990; Walsh et al., 1998; Rajan, 2000; Lauer and May, 2011; Liberman et al., 2014; Boero et al., 2018). The efferent system is also implicated in disorders such as auditory neuropathy where its function is diminished (Hood et al., 2003; Valero et al., 2018), and in tinnitus and hyperacusis where it is hyperactive (Knudson et al., 2014; Wilson et al., 2017; Wojtczak et al., 2017). The most caudal and widely investigated of these feedback mechanisms are the medial olivocochlear reflex (MOCR) and the middle ear muscle reflex (MEMR). The MOCR inhibits

cochlear amplification by limiting outer hair cell (OHC) motility (Siegel and Kim, 1982; Guinan and Gifford, 1988) and the MEMR reduces signal transfer through the middle ear by stiffening the ossicular chain (Borg et al., 1983; Liberman and Guinan, 1998). For decades, the MEMR has been clinically used to differentiate cochlear vs. neural pathologies (Jerger et al., 1974; Borg et al., 1983; Berlin et al., 2005). However, a reliable test of the MOCR currently does not exist. To fill this longstanding gap, here we describe a time-course and click-evoked otoacoustic emission (CEOAE)-based method that has the potential to serve as a simple and efficient test of efferent modulation of cochlear function.

Given that the MOC fibers directly innervate the OHCs (Warr and Guinan, 1979), OAEs provide a non-invasive means to investigate the influence of the MOCR on the OHCs (Guinan, 2006, 2014; Lopez-Poveda, 2018). Typically, a change in the OAE amplitude is monitored in the ipsilateral ear in response to MOCR activation in the contralateral ear with broadband noise (BBN; referred henceforth as the conventional method, see **Figure 1A**). While this method is convenient, it can be improved further in several ways:

(1) Conventional methods only test the contralateral pathway of the bilateral MOCR reflex system. Because BBN is presented in the contralateral ear to elicit the MOCR, no meaningful estimate of the MOCR is possible in this ear. If the ipsilateral or bilateral MOCR were to be estimated, forward masking techniques (Berlin et al., 1995; Boothalingam et al., 2018) or notched spectrum-noise methods (Backus and Guinan, 2006) must be employed. However, forward masking methods are time onerous and only capture the decaying segment of the MOCR (Backus and Guinan, 2006) and notched-spectrum-noise methods are not conducive for all types of OAEs (e.g., clicks). Estimating the ipsilateral and bilateral MOCR activity independently as well as in combination would simply provide a reductionist as well as holistic examination of the MOCR system (Guinan, 2014).

(2) In the conventional method, the stimulus itself can inadvertently activate the ipsilateral MOCR to unknown degrees, introducing uncertainties in MOCR magnitude estimation (Guinan et al., 2003; Boothalingam and Purcell, 2015; Boothalingam et al., 2018).

(3) Multiple MOCR studies have reported on the rather sub-par test-retest reliability of the conventional method (Mishra and Lutman, 2013; Stuart and Cobb, 2015; Mertes and Leek, 2016; Killan et al., 2017). This issue may, in part, be due to the reliance of the conventional method on "block averaging" (XYXY in **Figure 1A**) which is vulnerable to participant-related artifacts (e.g., change in middle ear pressure over time, probe drifts, etc.). The vulnerability comes from the temporal separation of the OAEs measured with and without the contralateral elicitor. This separation ranges between seconds to minutes across studies. Longer the gap between conditions, higher the risk of spurious changes in OAE level and probe drifts (Goodman et al., 2013).

(4) As illustrated in **Figure 1A**, conventional methods reduce the MOCR inhibition to a single data point in time, essentially decimating any data on reflex kinetics. That is, the evolution of the reflex over time cannot be gleaned from these methods.

As such, there is persistent uncertainty as to whether the change in OAE is due to the MOCR, participant-related artifact, or a systematic shift in measurement parameters.

Here, we propose a method that re-purposes the OAE-evoking clicks to also elicit and monitor MOCR activity. Click parameters used in this approach were identified in our prior work to optimally activate the MOCR while allowing adequate time for extracting CEOAEs (Boothalingam and Purcell, 2015; Boothalingam et al., 2018). Here, we extend the previous findings by employing these parameters [level: 80 dB peak-to-peak equivalent (pe)SPL; rate: 62.5 Hz] to test whether MOCR magnitude and time-constants can be extracted with either ear (left/right) and bilateral stimulation. This method is illustrated in **Figure 1B**.

The proposed method overcomes the limitations of the conventional method in the following ways:

(1) By using the same clicks that evoke CEOAEs to activate the MOCR, we relinquish the need to use a separate noise elicitor in the contralateral ear. This freedom from noise elicitor allows us to measure CEOAEs in both ears simultaneously and, consequentially, index the bilateral MOCR activity. Ironically, the limitation of this method is that the contralateral pathway cannot be evaluated separately.

(2) Because the proposed method does not require separate with- and without-noise conditions, conventional block-averaging is not necessary. As illustrated in **Figure 1B**, the entire click-train is averaged, which includes both the baseline (time zero) and the subsequent change in CEOAE over time. This short duration, unlike conventional methods

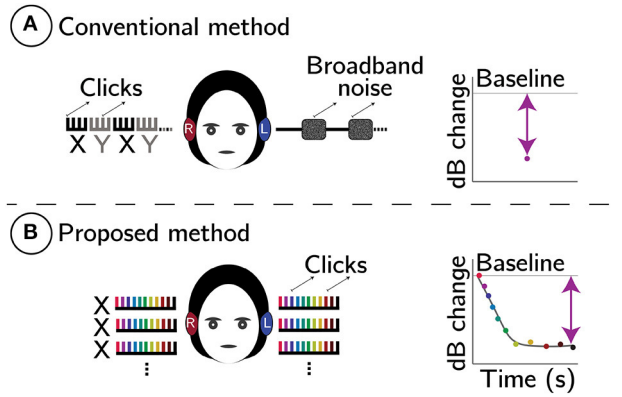

FIGURE 1 | Schematic comparison between conventional **(A)** and proposed **(B)** MOCR methods. In both panels, comb-like structures represent click trains. In **(A)**, X (black) are baseline CEOAEs without, and Y (gray) are with contralateral noise elicitor, respectively. Plots on the right are predicted change in CEOAE level (dB). In **(B)**, whole click-trains (X) are averaged and CEOAE inhibition is estimated using a two-term exponential fit to the change in CEOAE level, Δ, as a function of time. The line-fit also provides reflex kinetics (time-constants).

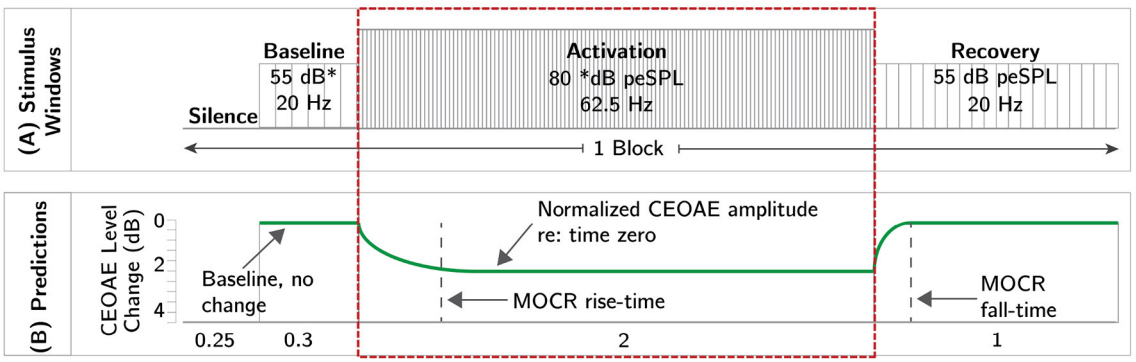

FIGURE 2 | Schematic of the experimental paradigm. Panel **(A)** illustrates the temporal order of different windows presented in the experiment. Panel **(B)** illustrates the predicted change in CEOAE level across different windows (green curves). The activation window, highlighted by the dotted red rectangle, is the same as the click paradigm described in **Figure 1B**. The duration of each click window are provided at the bottom of panel **(B)**.

where the with- and without-noise conditions are temporally separated, is predicted to minimize the undue influence of participant-related artifacts and/or probe drifts.

(3) In contrast to the uncertainty in the measured change in OAE being attributed to the MOCR in conventional methods, the well-established time-course of the MOCR—a two-term exponential with fast and slow time constants (Liberman et al., 1996; James et al., 2005; Backus and Guinan, 2006)—is exploited in the proposed method to determine if the change in the OAE is indeed due to the MOCR. While the time-course information can be obtained using contralateral noise in a conventional paradigm, it is time prohibitive and can only be obtained for one ear at a time.

We hypothesize that click-train averaging will preserve the time-course of the MOCR allowing for the extraction of MOCR magnitude and kinetics. The predicted change in CEOAE level is grossly illustrated in **Figure 1B** and more specifically in **Figure 2B**. The overarching goal of this work is to demonstrate the feasibility of the proposed CEOAE and time-course-based test of the MOCR.

2. METHODS

2.1. Participants

A total of 17 participants in the age range 18–30 yrs were recruited for the study in compliance with the guidelines of the Northwestern University Institutional Review Board and were compensated monetarily for their participation. All participants had an unremarkable otoscopic examination, hearing thresholds $\leq$20 dB HL between 0.25 and 8 kHz (Audio Traveler AA220, Interacoustics, Assens, Denmark) and clinically normal tympanometry (GSI TympStar, Grason-Stadler Inc., Eden Prairie, MN). Participants were also required to have measurable distortion product OAE (DPOAE) (2f1-f2; f2/f1 = 1.22; L1/L2 = 55/40 dB SPL) with signal-to-noise ratio >6 dB between 0.5 and 6 kHz. Middle ear muscle reflex (MEMR) thresholds evoked using clicks presented at 100 Hz were monitored using a 226 Hz tonal probe (GSI TympStar, Grason-Stadler Inc., Eden Prairie, MN) and were required to be >75 dB HL for inclusion. Two participants were rejected due to this inclusion criterion. One more participant was rejected due to the presence of more than five spontaneous OAEs (SOAEs) that were ~3 dB above the noise floor (Boothalingam et al., 2018). SOAEs influence CEOAE-based MOCR estimation as well as MEMR. Therefore, it is prudent to limit SOAE contributions. The final sample size was 14 [Mean age = 21.1; standard deviation (SD) = 1.9 yrs].

All testing was completed inside a double-walled sound booth where participants sat in a comfortable recliner for the duration of the experiment. During the experiment, participants watched a silent closed-captioned movie of their choice. Participants were encouraged to relax, not swallow, and stay awake during periods of stimulus presentation. Breaks were provided every 8 min, during which participants were encouraged to stretch, drink water, and do other noisy activities that were discouraged during recording periods. Throughout the experiment, OAE probes in both ears of participants were sealed using earmold putty to avoid slippage/drifts. The entire experiment took roughly 3 hrs to complete per participant. All testing was completed in a single session.

2.2. Stimulus Generation

All stimuli were digitally generated in MATLAB (v2016b; Mathworks, MA, USA) at a sampling rate of 96 kHz and a bit depth of 24. Similar to Boothalingam et al. (2018), clicks were generated in the frequency domain using a recursive exponential filter (Zweig and Shera, 1995; Charaziak et al., 2020) for band-limiting the click between 0.8 and 6 kHz (~108 μs long). Bandpass clicks were used to focus the stimulus energy in the frequency regions where the MOCR is most prominent (Lilaonitkul and Guinan, 2012; Zhao and Dhar, 2012). In addition, bandpass clicks produced less loudspeaker ringing in our set-up compared to a single sample impulse.

2.3. Instrumentation and Calibration

Instrumentation was similar to that described in Boothalingam et al. (2018). Briefly, signal delivery and data acquisition

were controlled through an iMac computer (Apple, CA, USA) running Auditory Research Lab Audio Software (ARLas v4.2017 Goodman, 2017) on MATLAB at 96 kHz. Digital-to-analog and analog-to-digital conversions were handled by an external sound card (Fireface UCX; RME, Germany) connected to the iMac via Firewire. Signals were delivered bilaterally via two ER2 insert receivers coupled to two separate ER10B+ probe assemblies (Etymotic Research, IL, USA). The second port in the ER10B+ probe was coupled to identical dummy loudspeakers bilaterally. Ear canal pressure was registered and amplified (+20 dB) by the ER10B+ probe microphone and pre-amplifier, respectively.

The root-mean-square (RMS) level of BBN was calibrated in a Zwislocki ear simulator. Click levels were also first calibrated in a Zwislocki ear simulator where its peak-to-peak amplitude was matched with a 1 kHz sine tone. In addition, an in-ear calibration was performed in each participant. In this approach, a sample of clicks were played in the participant's ears before the start of every condition and any deviations from the expected peSPL at the probe-tip were corrected.

2.4. Experimental Paradigm

A schematic of the experimental paradigm is illustrated in **Figure 2**. The hierarchy of terminologies is as follows. Each click presentation and the silent duration which follows until the onset of the next click is an "epoch." Therefore, epoch durations vary with click rate. Clicks with different levels and rates served different purposes and were grouped into "windows." The difference in column width in each window in **Figure 2A** represents the epoch duration and the height represents the click level. Four different "windows" made up a single "block." Blocks were repeated 500 times. The "silence" window (250 ms), where no stimulus was presented, allowed the MOCR to return to baseline functioning (Backus and Guinan, 2006). In the "baseline" (300 ms), low-level (55 dB peSPL) and slow-rate (20 Hz) clicks that are known to not activate the MOCR or the MEMR were presented (Boothalingam and Purcell, 2015; Boothalingam et al., 2018). CEOAEs in the baseline window served as confirmation for MOCR activity starting from the baseline no activity in the "activation" window where higher level (80 dB peSPL) and faster rate (62.5 Hz) clicks were presented for 2 s. The click level and rate used in this window are based on our prior work that demonstrated robust MOCR activation with little-to-no evidence of MEMR activation (Boothalingam and Purcell, 2015; Boothalingam et al., 2018). Finally, the same slow rate and low level clicks from the baseline were presented again for 1 s in the "recovery" window to capture the MOCR decay. The same paradigm was presented in three lateralities which included two unilateral stimulations, left- and right-only stimulation, and one bilateral stimulation. However, for the sake of brevity, ear canal recordings from only one ear from bilateral stimulation is discussed in this paper. This includes an equal number (7) of right and left ears.

Note that because the click levels and rates are different across windows, the evoked OAEs cannot be considered as a continuous function over time. However, despite the rate/level differences between activation and recovery windows, the elicited MOCR activity is considered a continuous function of time across these two windows. This is because it is the MOCR elicited by the click-train in the activation window that is being captured in the recovery window. Slowing the click rate and lowering the click level is essential in this process because continuing the same high rate and level from the activation window will not allow the MOCR to decay.

2.5. CEOAE Extraction

Raw microphone pressure recordings were processed offline using custom scripts written in MATLAB. First, all pressure recordings were bandpass filtered between 0.8 and 4.2 kHz—close to the bandpass frequency of the click stimulus. Next, any epochs that had an RMS amplitude >2.25 times the interquartile range (within-participant) were rejected as containing artifacts. Overall, less than 10% of the data were rejected across included participants.

All MOCR analyses were conducted on CEOAEs time-windowed between 4.5 and 15 ms relative to time zero. Time zero was defined as the location of the peak of the click stimulus. Hann ramps (1 ms long) were applied at the start and end of the CEOAE waveform. Prior to any analysis, epochs within the different stimulus windows were sub-averaged by a factor of 2. For instance, in the 125 clicks (62.5 Hz $\times$ 2 s) recorded per block in the activation window, adjacent epochs were averaged. This sub-averaging, while reducing the resolution of the time-course by a factor of 2 (32 ms instead of 16 ms in the activation window), allowed for estimation of CEOAEs at each time point from 1,000 epochs [500 repetitions $\times$ 2; Boothalingam and Goodman, 2021]. This step allowed us to reduce test times by half while still maintaining the quality of the recorded CEOAEs. The 32 ms resolution is smaller than the rise- and fall-time of the MOCR (Kim et al., 2001; Backus and Guinan, 2006). Therefore, this sub-averaging should not affect the quality of the time constants obtained.

Next, within each epoch, CEOAEs were considered in the time-frequency domain to more precisely extract the signal of interest. A time-frequency representation of the OAE was constructed using a bank of overlapping gammatone filters with center frequencies between 0.8 and 4.2 kHz (Goodman et al., 2021). The filters were based on models of human auditory filters (Glasberg and Moore, 1990). Stimulus frequency (SF) OAE-based delays (Shera et al., 2002, 2008) were used to time window the filtered waveforms so as to only include CEOAEs within the expected delays ($\pm$ 20%) at each frequency. After time windowing, the filtered waveforms were added back together to yield composite waveforms. This approach improves signal-to-noise ratio because noise energy is excluded from temporal regions where no OAEs are expected in each filer band. As such, the steps described herein allowed for the extraction of time (within each epoch) and frequency (specific bands of interest)-based CEOAEs at each time point in the respective windows.

Following extraction, the CEOAEs within each frequency band were averaged across time (within each epoch). The averaging process also included taking the energy-weighted average within each frequency band. That is, the spectral energy, which is the square of the pressure magnitude at each Fourier frequency, was used as weights for computing the weighted mean

across frequency. This reduced the contribution of frequencies with small OAE magnitude relative to frequencies with higher OAE magnitude. Averaging was performed separately for each point in the time series, i.e., every 32 ms. This process reduced the time-frequency representation of the CEOAE to 7 spectral magnitudes corresponding to the passband frequencies of a bank of third-octave filters with nominal center frequencies 1, 1.2, 1.6, 2, 2.5, 3.2, and 4 kHz, where the MOCR effects are predominant (Lilaonitkul and Guinan, 2012; Zhao and Dhar, 2012). The CEOAE magnitude at each frequency and time point was calculated as the magnitude of the mean across the 500 repetitions. Within each frequency band, noise floor was estimated as the standard error of the mean of the CEOAE spectra (Goodman et al., 2009). An SNR criterion of 12 dB was imposed for CEOAEs at each frequency to be included. Frequencies, typically spectral notches, where the SNR was lower than 12 dB were not included in the averaging process.

2.6. Time-Course Analysis for MOCR Estimation

For each frequency band, the magnitude in Pascals at each point in the time series was divided by the magnitude in Pascals at the first time point for the activation window and the last time point for the recovery windows. Recall that each time point is an average of 1,000 click epochs across 0-32 ms, i.e., two consecutive epochs in time repeated 500 times. This use of within-window baseline is one of the strengths of the proposed method for MOCR estimation as it does not require a separate baseline measurement. Referred to hereafter as Δ, this metric of relative change was then expressed in dB. This final step allowed for easier visualization of the change in CEOAE over time across frequencies and participants. As illustrated in **Figure 2B**, no change in the CEOAE over time can be imagined as a straight line at 0 dB. A negative Δ, i.e., reduction in CEOAE magnitude, would indicate potential MOCR activation.

MOCR activation was quantified as the change in Δ at 2 s, the final time point in the activation window, and termed Δ_{max}. The change and the associated rise- and fall-times, were estimated using a two-term exponential line fit to the CEOAE data, similar to the implementation of this method for MEMR estimation (Boothalingam and Goodman, 2021) and based on DPOAE rapid adaptation (Liberman et al., 1996; Kim et al., 2001; Srinivasan et al., 2012). A two-term exponential fit has previously been shown to provide a good estimate of the MOCR as the MOCR activation works on at least two time scales: fast and slow (Sridhar et al., 1995; Liberman et al., 1996; Kim et al., 2001; Backus and Guinan, 2006). The two-term exponential was of the form:

$$f(t) = C + m_f * e^{(-t/tau_f)} + m_s * e^{(-t/tau_s)}, \quad (1)$$

where f is the fit as a function of time, t. The variables m_f and m_s are the magnitude of the fast and slow components of the fits, respectively. The variables tau_f and tau_s are the fast and slow time-constants, respectively. C is a constant term representing offset along the y axis. To determine if the Δ approximated by the two-term exponential fit is statistically significant, we employed a permutation-based implementation of the Heller-Heller-Gorfine (HHG; Heller et al., 2013) test as described by Boothalingam and Goodman (2021) for MEMR estimation. Briefly, the fit and Δ were compared as two vectors hypothesized to have no association, i.e., at least one of the two vectors, more likely the Δ, changes randomly over time. Significance of the comparison (p-value) was obtained by generating confidence intervals from bootstrapping the HHG test 1,000 times. Because seven such tests were conducted for any given laterality/window, the p-values were corrected for performing multiple comparisons using Bonferroni correction. A significant fit was considered as MOCR activation. Our pilot data indicated a lack of MOCR activation in the baseline window. Therefore, we regressed Δ in the baseline window against time using simple linear regression to test for any systematic change over time.

2.7. Test for MEMR Activation

The presence of MEMR may influence MOCR activation, and therefore the recorded responses must be carefully examined. Activation of the MEMR alters the impedance characteristics of the middle ear. Because impedance is frequency dependent, it is important to recognize that different frequencies are affected differently. At frequencies below ∼0.8 kHz and above ∼1.5 kHz, there is an increase in stimulus reflectance whereas there is a reduction in reflectance between ∼0.8 and 1.5 kHz (Feeney and Keefe, 1999; Feeney et al., 2017; Boothalingam and Goodman, 2021). We used the same time-course method used for the MOCR, except the use of time-frequency analysis, and as described by Boothalingam and Goodman (2021), to determine MEMR activation in all seven frequency bands. The difference between the MOCR and the MEMR analyses is that the stimulus waveform (0–4 ms) was analyzed to determine the presence of the MEMR while the CEOAE waveform was analyzed to determine the presence of the MOCR.

3. RESULTS

3.1. MEMR Activation Has Minimal Effect on MOCR Magnitude

While our previous broadband, and arguably less sensitive, approach to MEMR detection (Boothalingam et al., 2018) suggested that 80 dB peSPL clicks presented at 62.5 Hz should not significantly activate the MEMR, studying stimulus Δ in narrow bands of frequencies in the present approach shows MEMR activation in 100% of the participants for all three lateralities. Representative data for bilateral stimulation from two participants, one with large and one with small MEMR activation, are presented in **Figure 3**. Notice that although both participants demonstrate statistically significant stimulus Δ in the activation window, their Δ_{max} are vastly different, especially at the lower frequencies. For instance, Δ_{max} in the 1.3 kHz band for n13 is 1.8 dB compared to 0.0026 dB for n5. That is, the Δ_{max} of MEMR for n13 is ∼700 times larger than that of n5. A natural question then is: do all MEMR activations necessitate influence on MOCR Δ_{max}? To answer this question, we computed Pearson correlation coefficients between MOCR and MEMR Δ_{max} at and across all seven bands of frequencies.

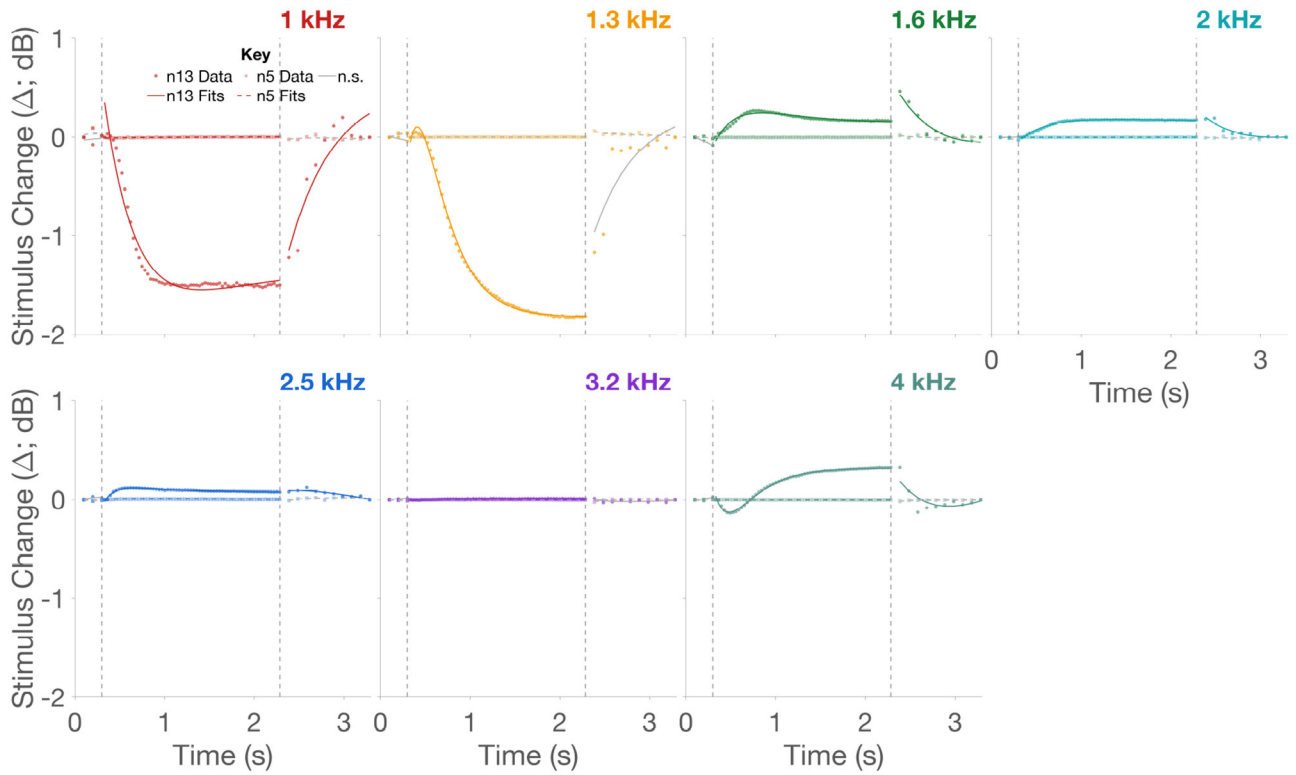

FIGURE 3 | Stimulus change over time in two representative participants. Panels separate 1/3rd octave-band frequencies. Time-course data for the two participants, n5 and n13, are shown as squares and circles, respectively. Fits to the data are shown as dashed and unbroken lines, respectively. Colors represent the different frequencies. Fits lines in the respective color are statistically significant while fit lines in gray are not. The two vertical dashed lines in panels indicate the temporal separation between the three windows.

Scatter plots of absolute MOCR Δ_{max}, i.e., MOCR magnitude vs. absolute MEMR Δ_{max}, i.e., MEMR magnitude, for 1, 2, and 4 kHz bands are plotted in **Figure 4**. Only three of the seven frequencies are shown for brevity. As seen in the scatter plots, MEMR magnitude only correlates with MOCR magnitude when the outlier (n13) is included. Despite including the outlier, correlations were only significant for MOCR magnitude at 1 and 1.3 (not shown) kHz. Revaluation without the outlier did not produce any significant correlations even before correcting alpha for performing multiple comparisons using the False Discovery Rate (FDR) method (Benjamini and Hochberg, 1995). Evidently, the large change in stimulus in this particular participant, likely due to MEMR activation, has had an impact on their MOCR magnitude. As a group, however, the lack of correlations between MEMR and MOCR magnitude suggest that small changes in stimulus level, even if they are statistically significant, do not significantly affect MOCR estimates, at least in the 1–4 kHz frequency range. Therefore, only the n13 participant data were excluded from all remaining analyses. As such, the changes in CEOAE magnitude reported here are likely predominantly driven by the MOCR, not the MEMR.

3.2. Clicks Elicit Robust CEOAE Inhibition

Mean Δ for all three windows across frequencies and lateralities are shown in **Figure 5**. As predicted in **Figure 2B**, there is no significant activity in the baseline window. This is followed by a significant Δ in the activation window, and finally the Δ returns to baseline in the recovery window. That is, clicks presented at 80 dB peSPL and 62.5 Hz (activation window) produced a significant inhibition of CEOAEs over the 2 s period in all three lateralities: right, left, and bilateral stimulation. When averaged across the seven frequencies, 91.2% of the two-term exponential fits (from a total of 91; 13 participants x 7 frequencies) were significant in the activation window for bilateral stimulation compared to 27.5% in the recovery window. The lower number of significant fits in the recovery window is likely due to lower click level (55 dB peSPL) and coarser time resolution (100 ms). For the right and left ear-only stimulation, the number of significant MOCR activations were lower at 58.2 and 71.4%, respectively. For a better comparison across lateralities, only the fits are presented in **Figure 6**. Similarly, for the recovery window, the percentage of significant fits to data for the right and left ear-only stimulation were also lower at 10.9 and 20.9%, respectively. No fits (0%) in the baseline window were significant for any of the three lateralities.

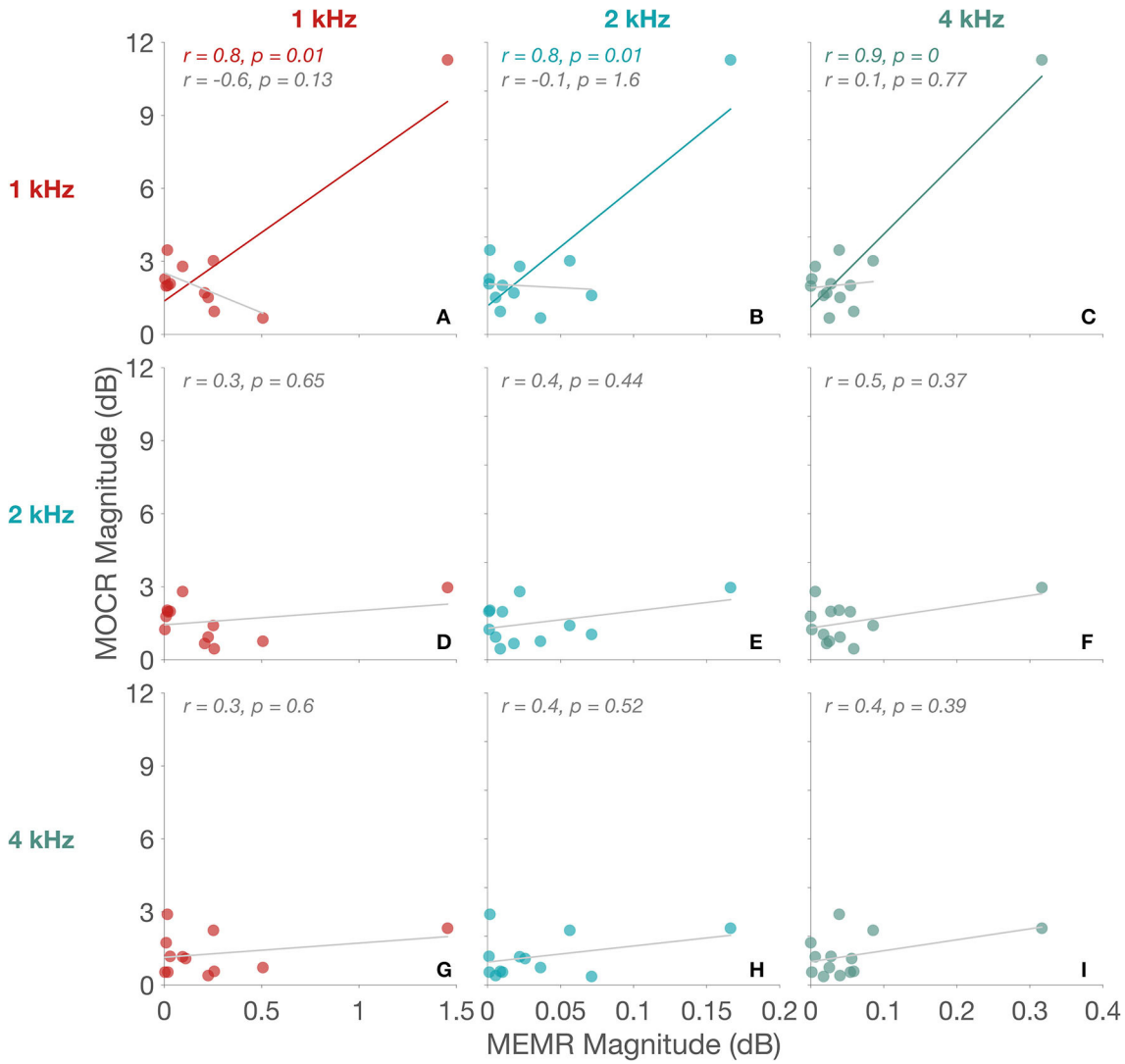

FIGURE 4 | MEMR vs. MOCR. Panels **(A–I)** separate 1/3rd octave-band frequencies. MEMR magnitude are along the x-axis and MOCR magnitude are along the y-axis in all panels. MEMR frequencies are differentiated in panel columns and MOCR frequencies in panel rows. Comparison frequencies in each panel are the two frequencies intersecting the specific panel [e.g., panel **(A)** compares both MOCR and MEMR magnitude at 1 kHz]. Significant fits are indicated by frequency specific colors and non-significant fits are in gray. Corresponding Pearson correlation coefficient (*r*) and the *p*-value are provided at the top of each panel.

Note that the mean Δ and all further analyses on MOCR/MEMR Δ_{max} and time constants were conducted only on data with significant fits to Δ. Due to the small number of significant fits in the recovery window and no significant fits in the baseline no further inferential statistics were conducted for data from these two windows.

MOCR Δ_{max} in the activation window extracted from fits to Δ are plotted as box plots in **Figure 7**. Two crucial observations can be made from **Figures 5**–7. (1) As expected based on binaural integration, bilateral stimulation produced larger, more than twice the MOCR inhibition (1.69 ± 1.2 dB; ± 1SD) relative to right (0.61 ± 0.4 dB) and left (0.62 ± 0.4 dB) ear-only stimulation. This is consistent with several prior reports (Berlin et al., 1995; Backus and Guinan, 2007; Lilaonitkul and Guinan, 2009a; Boothalingam et al., 2018, 2019). (2) At least for the bilateral stimulation, not all frequencies (**Figure 6**; Bilateral Panel) appear to be inhibited to the same extent. The largest mean inhibition is observed at the lower frequencies with inhibition progressively getting smaller. This is also consistent with prior work showing smaller MOCR activation above ∼3 kHz (Goodman et al., 2013).

To study the data inferentially, a linear mixed-effects model was used. Laterality and frequency were fixed-effects while MOCR Δ_{max} was the dependent variable with random intercepts for each participant. An analysis of variance (ANOVA) of the model suggested a significant interaction between frequency and laterality [$F_{(12, 180)} = 1.9$, $p = 0.04$] with significant main effects of both laterality [$F_{(2, 180)} = 35.3$, $p < 0.001$] and frequency [$F_{(6, 180)} = 5.9$, $p < 0.001$]. *Post-hoc* pairwise comparisons for

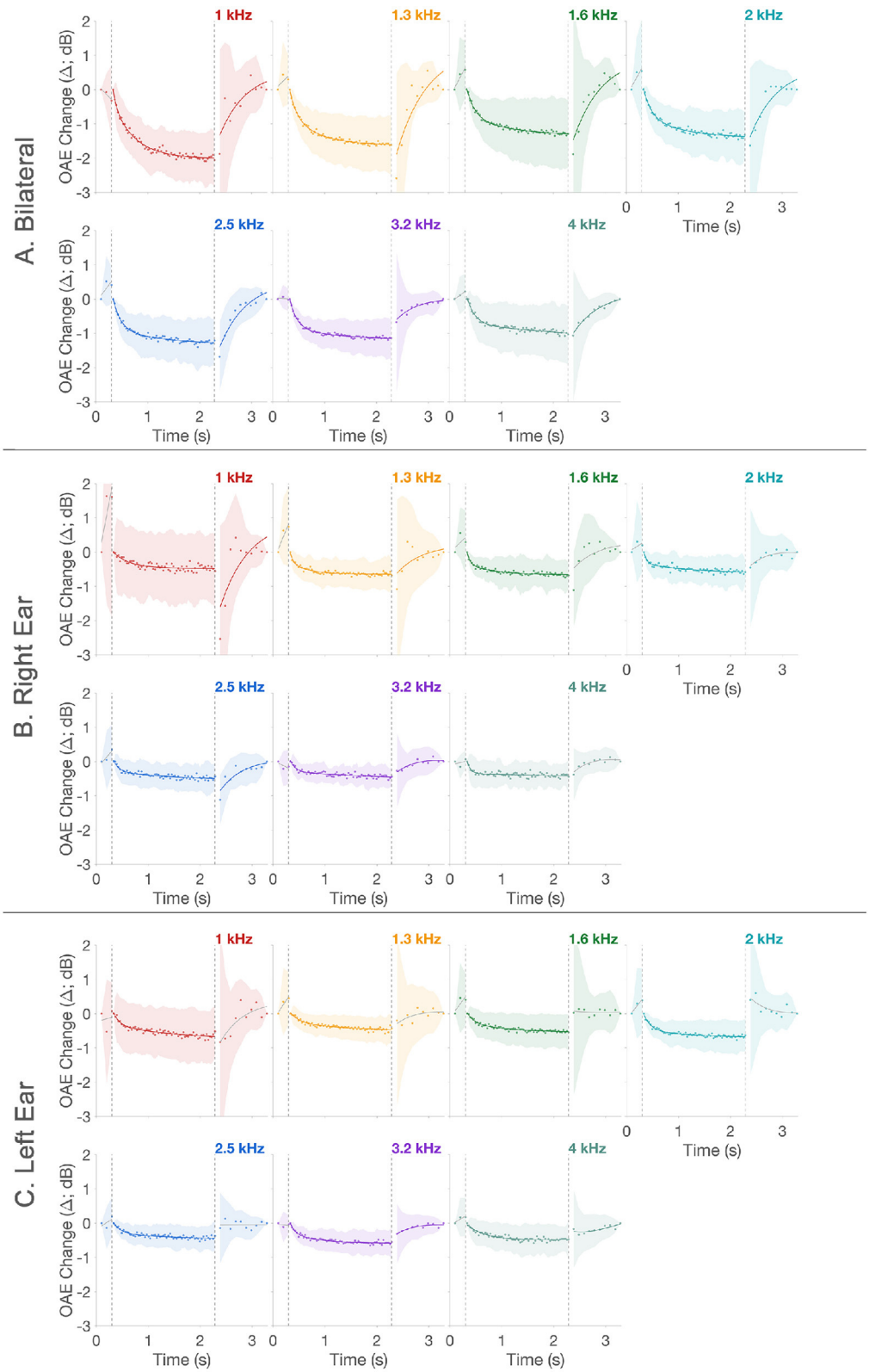

FIGURE 5 | CEOAE change over time. Panels **(A–C)** show bilateral, right, and left ear mean time-course data, respectively. In all panels, scatter plot is the mean CEOAE change (Δ) across all 13 participants. The lines are statistical model fit to the data, linear (baseline) and two-term exponential (activation and recovery). Shaded region around the data represents ± 1SD around the mean. Colors represent the different frequencies. Fits lines in the respective color are statistically significant while fit lines in gray are not. The two vertical dashed lines in panels indicate the temporal separation between the three windows.

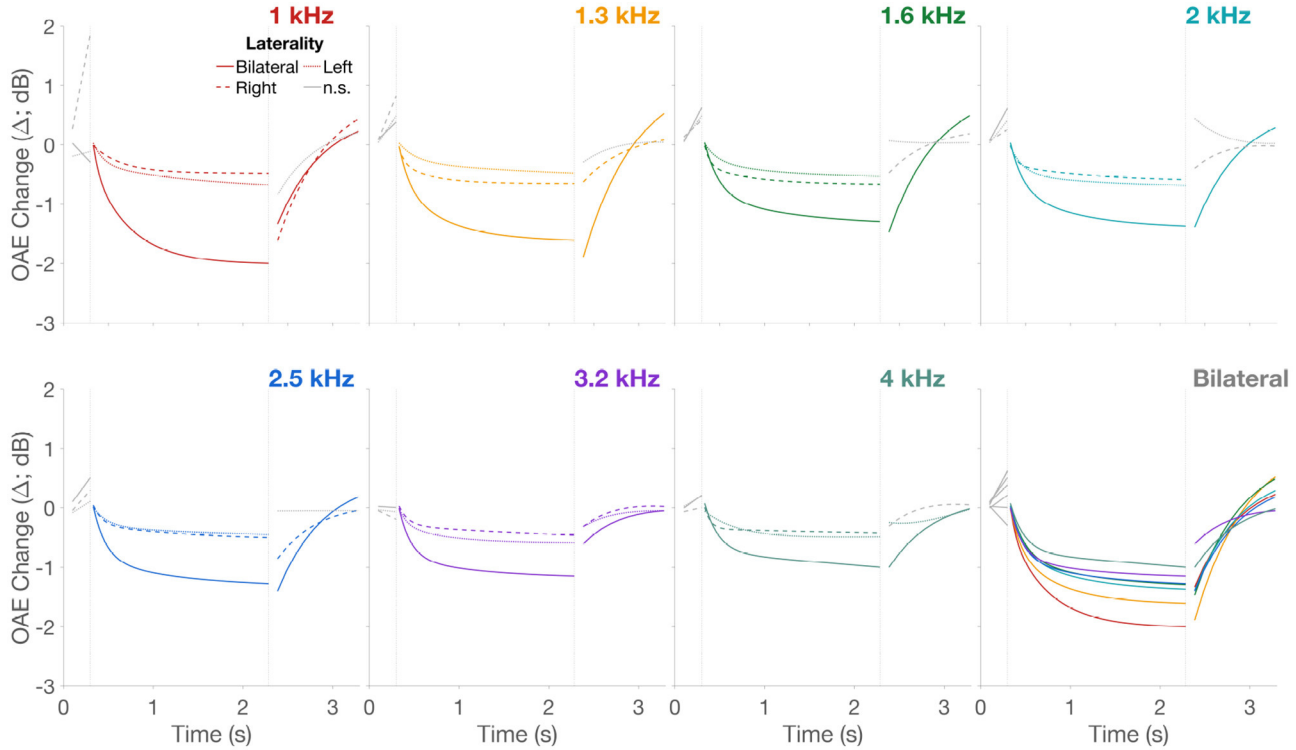

FIGURE 6 | CEOAE change over time. Only the fits from **Figure 5** are plotted to allow direct comparison across the three lateralities. Unbroken lines are bilateral, dashed lines are right, and dotted lines are left ear stimulation, respectively. Panels are separated by frequency in addition to the different colors for the 7 frequencies. Lines in gray are non-significant fits. Fits across all frequencies, only for the bilateral stimulation, is presented in the last panel "Bilateral" to allow direct visual comparison of CEOAE Δ as a function of frequency.

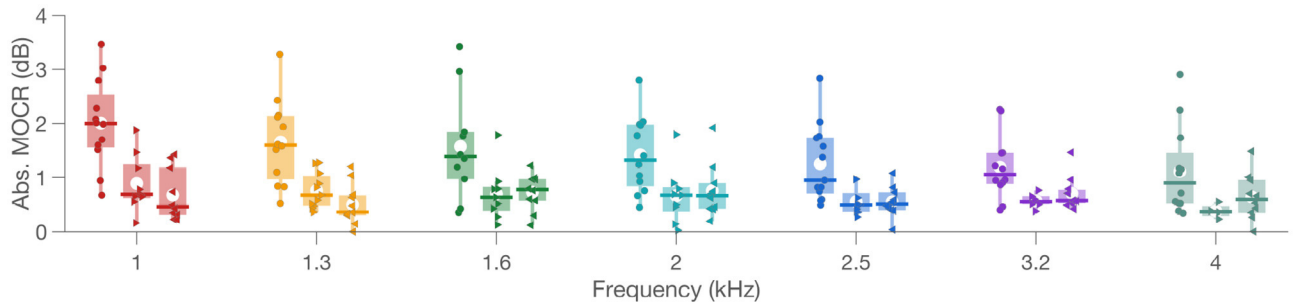

FIGURE 7 | MOCR magnitude. Box plots show individual MOCR magnitude, i.e., absolute Δ_{max}, as filled colored shapes at respective frequencies along the x-axis. Circle is bilateral, right-pointing triangle is right ear, and left-pointing triangle is for left ear data. Colors represent frequency in the x axis. In the box plots, the box represents the interquartile range, white circle is the mean, vertical line is the data range, and the horizontal line is the median.

frequency differences within each laterality were conducted using *t*-tests corrected for multiple comparisons using the FDR method. Only three comparisons (out of 21) were statistically significant in the bilateral condition [1 vs. 2 kHz ($p = 0.035$); 1 vs. 2.5 kHz ($p = 0.038$); 1 vs. 4 kHz ($p = 0.009$)]. *Post-hocs* for the laterality effect suggested significantly larger MOCR Δ_{max} in the bilateral compared to both left [$t(12) = -8.1$; $p < 0.001$] and right [$t(12) = -5.8$; $p < 0.001$] ear stimulations, as expected. However, left and right ears were not significantly different [$t(12) = -0.65$; $p = 0.53$].

3.3. Click-Elicited CEOAE Inhibition Follows a Physiological Time-Course

Time constants derived from the two-term exponential fits are shown in **Figure 8**. The mean fast rise time (tau_f) of the MOCR (averaged across frequencies) for the three lateralities

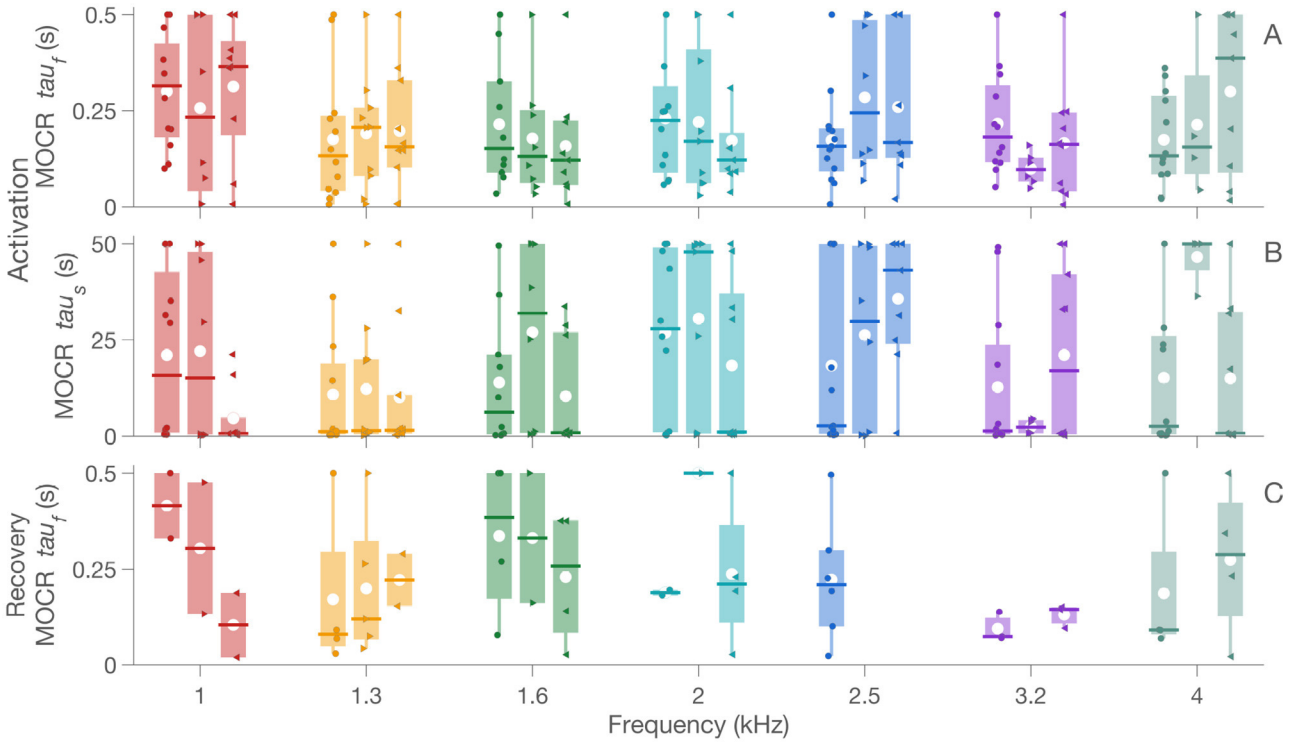

FIGURE 8 | MOCR *tau*. Panels **(A,B)** show data for tau_f and tau_s in the activation window and panel **(C)** shows data for tau_f in the recovery window. Similar to **Figure 7**, box plots show individual MOCR *tau* as colored boxes at respective frequencies along the x-axis. Colored circle is bilateral, right-pointing triangle is right ear, and left-pointing triangle is for left ear data. In the box plots, the box represents the interquartile range, white circle is the mean, vertical line is the data range, and the horizontal line is the median.

were essentially the same: 0.22 ± 0.15, 0.21 ± 0.16, and 0.21 ± 0.17 s for bilateral, right, and left ear stimulation, respectively. These average values are consistent with the 0.28 ± 62 s reported by Backus and Guinan (2006). Unlike MOCR Δ_{max}, there was no effect of laterality for the rise time tau_f [$F_{(2,198)} = 0.01$, $p = 0.99$], frequency [$F_{(6,198)} = 1.0$, $p = 0.39$], or their interaction [$F_{(12,198)} = 0.9$, $p = 0.46$]. This result is also consistent with the findings of Backus and Guinan (2006) where they demonstrated the independence of MOCR time constants from elicitor level or laterality effects.

For the slow rise time, tau_s, however, time constants for the three lateralities were slightly more variable: 16.8 ± 20, 23.8 ± 16.9, and 17 ± 17.5 s for bilateral, right, and left ear stimulation, respectively. Nonetheless, these values are also similar to those reported by Backus and Guinan (2006). The mixed-effects model suggested a significant interaction between laterality and frequency [$F_{(12,198)} = 2.2$, $p = 0.02$] and a main effect of laterality [$F_{(2,198)} = 3.3$, $p = 0.04$]. The fixed-effect of frequency was not significant [$F_{(6,198)} = 0.97$, $p = 0.44$]. Because the main effects of frequency was not significant, data were collapsed across frequency to test for the effect of laterality. This *post-hoc* analysis, with *p*-values corrected for multiple comparisons using the FDR method, suggested no difference between right and left [$t(12) = 1.1$; $p = 0.44$], or left and bilateral [$t(12) = -0.1$; $p = 0.92$], or right and bilateral [$t(12) = -1.5$; $p = 0.45$].

The mean fall times (averaged across frequencies) were 0.22 ± 0.13, 0.33 ± 0.17, and 0.19 ± 0.14 s for bilateral, right, and left ear stimulation, respectively. These values are slightly longer than the 0.16 ± 0.5 s reported by Backus and Guinan (2006). Although no statistics were performed for the fall times (recovery window) due to the sparseness in the data, raw fall tau_f are shown in **Figure 8**. These values must be interpreted cautiously as only between 10 and 27% of the data produced significant fits.

To test whether the MOCR Δ_{max} at 2 s is statistically different from that at earlier times (1, 1.25, 1.5 s) we performed *t*-test, corrected for multiple comparisons using the FDR method. Four of the 7 frequencies (1, 1.3, 2, and 2.5 kHz; p <0.022) at 1 s was significantly different from that at 2 s. At 1.25 s, this number reduced slightly to 3 of 7 frequencies (1.3, 2, and 2.5 kHz; p <0.03). At, 1.5 s, only one of the 7 frequencies (2 kHz; $p = 0.035$) produced significantly different MOCR Δ_{max} than that at 2 s. This result suggests that click train duration between 1 and 1.5 should be sufficient to estimate the MOCR using the proposed approach.

4. DISCUSSION

The results presented here provide evidence that clicks can be used to elicit and estimate MOCR activity simultaneously without the need for a contralateral noise elicitor.

4.1. MOCR Magnitude

The click level (80 dB peSPL) used in the present approach is higher in comparison to conventional MOCR methods (typical: 55–75 dB peSPL Veuillet et al., 1991; Hood et al., 1996; Goodman et al., 2013; Lewis, 2019). The 80 dB peSPL was chosen for its ability to elicit robust MOCR activity in a forward masking paradigm where the MOCR monitoring clicks were presented at 55 dB peSPL (Boothalingam et al., 2018). Higher click levels capture less MOCR activity as there is relatively less cochlear amplification for the MOCR to inhibit (Goodman et al., 2021). However, because our approach uses the same clicks to both elicit and monitor MOCR activity, the chosen click level must be able to play both roles. As such, the 80 dB peSPL is a compromise between (1) activating adequate MOCR activity (higher levels preferred), (2) capturing maximum possible MOCR activity using OAEs (lower levels preferred), and (3) avoiding MEMR activation (lower levels preferred). The mean MOCR Δ_{max} in the bilateral condition (1.69 dB) is commensurate with the 1–2 dB OAE inhibition typically reported in conventional noise-based studies for both contralateral and bilateral stimulations. While the bilateral stimulation in the present study would be expected to produce larger MOCR activation than the conventional contralateral noise stimulation paradigm, it should be noted that noise is a more potent elicitor than clicks (Veuillet et al., 1991; Guinan et al., 2003). It thus appears that the reduced potency of clicks in eliciting the MOCR is offset by bilateral stimulation. Similarly, the reduced potency of clicks in the current paradigm is offset by capturing the MOCR activity at its temporal peak unlike forward masked bilateral paradigms that although use noise but capture only the decaying portion of the MOCR. Finally, Lewis (2019) demonstrated that the larger OAE SNR counteracts the reduction in cochlear amplification at higher stimulus levels by allowing better detectability of MOCR activation. Therefore, levels around 80 dB peSPL seem ideal for the present approach. Taken together, it can be argued that our stimulus choice accomplishes both activation and monitoring of the MOCR similar to currently available methods.

The level that was optimal for bilateral stimulation did not elicit adequate MOCR activity in the unilateral conditions. The mean MOCR Δ_{max} for both right and left ears were <1 dB. This result is not unexpected based on the known physiology of the MOC neurons in the brainstem. Bilateral stimulation activates both ipsilateral and contralateral MOC neurons in addition to binaural MOC neurons (Liberman and Brown, 1986; Liberman, 1988). Evidently, there is a considerable increase in the number of neurons that are activated during bilateral stimulation. Furthermore, bilateral stimulation also allows for the capture of both crossed and uncrossed MOC fiber action in the cochlea. Therefore, bilateral stimulation not only elicits larger activity but also provides a complete picture of the MOCR function by activating all types of MOC neurons and pathways. There is also a considerable inter-species difference in the distribution of contralateral vs. ipsilateral MOC neurons. For instance, about 90% of the neurons respond to ipsilateral sound (Liberman and Brown, 1986) in cats, while about 50–55% respond to ipsilateral sound (Robertson and Gummer, 1985; Brown, 1989) in guinea pigs. Although this distribution is unknown in humans, it can be surmised from OAE-based studies that there may not be a large difference between ipsilateral and contralateral neuron count as they produce similar MOCR magnitude (Guinan, 2006). However, ipsilateral and contralateral stimulations do indeed produce varying degrees of activation when narrowband stimuli are employed, thought to be driven by unknown central processes rather than the MOC neurons themselves (Lilaonitkul and Guinan, 2009a).

The results of *post-hoc t*-tests for MOCR Δ_{max} corroborate the larger MOCR activation at around 1 kHz consistently reported in the literature (Lilaonitkul and Guinan, 2012; Zhao and Dhar, 2012). The smaller MOCR Δ_{max} in the unilateral stimulation presumably was not large enough to demonstrate such frequency effects. As such, these results indirectly highlight the importance of the size of the MOCR magnitude to reliably study the influence of experimental variables on MOCR activity (e.g., task difficulty, attention). If the MOCR Δ_{max} were not large enough to capture the effects of such variables, the presence of a potentially underlying effect may be rejected in error. Taken together, it does appear that bilateral stimulation is a better approach to study the function of the MOCR more completely and robustly. Because bilateral stimulation, using forward masking, in conventional noise elicitor-based methods capture only the decaying portion of the reflex, the proposed time-course-based method provides a feasible solution for both research and clinic. It should, however, be noted that despite capturing only the decaying portion of the reflex, noise elicitor-based forward masking paradigms do produce MOCR magnitude comparable to the current approach. This is likely due to noise elicitors being more potent than clicks (Veuillet et al., 1991; Guinan et al., 2003).

A byproduct of measuring ipsilateral and bilateral stimulation is the ability to study binaural interaction. The larger bilateral MOCR (1.69 dB) relative to the sum of right and left ear MOCR (0.61 + 0.62 = 1.23 dB) demonstrates "binaural interaction" reported in our prior work that used forward masking (Boothalingam et al., 2016). While measuring binaural integration was not one of the motivations of this study or approach, observing such well-known effects in our method provides confidence, that the CEOAE inhibition observed here is quite likely driven by MOCR activation. Another aspect of the MOCR that can be readily compared in this approach is the difference between left and right ears. This difference can be studied using unilateral or bilateral stimulation. For instance, the results of *post-hoc* tests between unilateral left/right ear stimulation suggested no difference between left and right ears. Although this is contrary to some studies (Khalfa et al., 1997; Morlet et al., 1999; Bidelman and Bhagat, 2015), others have reported similar results (Philibert et al., 1998; Xing and Gong, 2017). While animal (Gifford and Guinan, 1987) and human (Backus and Guinan, 2007; Lilaonitkul and Guinan, 2009a,b) studies have shown similarities and differences in the effects of crossed vs. uncrossed fibers on OAEs, the ear asymmetry in MOCR function remains unsettled. Using bilateral and unilateral stimulations in the present method, this question could be explored further in future studies.

4.2. MOCR Kinetics

4.2.1. MOCR Activation

As predicted (**Figures 1**, **2A**), the CEOAE Δ, i.e., the MOCR demonstrated a rise and a fall time. The two-term exponential fit approximated the data well, and as a result, we were able to extract the fast, tau_f, and slow, tau_s, rise time constants. Further, the lack of any significant fits in the baseline suggests that any Δ in the baseline window are likely a result of fluctuations in background noise over time. As seen in **Figures 5**, **6**, these Δ can sometimes be larger than the Δ in the activation window. This is likely because CEOAEs elicited by the lower level clicks are less robust to background noise and therefore vary more over time. Crucially, these random fluctuations suggest that neither MOCR nor MEMR was active in the baseline window. Therefore, the MOCR and MEMR activity in the activation window can be surmised to have always started from the baseline, i.e., no activity, in every block. Alternatively, the non-significant Δ in the baseline window could be a result of relatively larger variance in this window, compared to the activation window, which may obscure small MOCR and MEMR effects.

The average tau_f across the three lateralities, 0.21 s, is commensurate with prior reports in humans (Backus and Guinan, 2006) as well as in animal models (Warren and Liberman, 1989; Liberman et al., 1996). Backus and Guinan (2006) also reported a slow time constant that was on the order of 10s of seconds and a medium time constant that was on the order of a few hundreds of milliseconds. An almost equal number of fits (participants × frequencies; 13 × 7) have tau_s of few hundreds of milliseconds (34%) and 10s of seconds (40%) in our data. It thus appears that both tau_f and tau_s in the present study may be mixtures of fast and medium, and medium and slow, time constants, respectively. We did not differentiate the *tau* into further smaller quantities as this was not the focus of the study. However, the corroboration with prior studies suggests that the time-course data reported here is of physiological, specifically of the MOCR, in origin. Prior reports have suggested an onset delay of the MOCR to be roughly between 25 and 60 ms (James et al., 2005; Backus and Guinan, 2006). The resolution of the time course in our approach, 32 ms, is too coarse to estimate such a short delay in the present study. Future studies that use clicks/tonebursts presented at faster rates (>62.5 Hz) when possible may be able to capture this detail with greater precision.

Also corroborating prior results (Backus and Guinan, 2006) were the lack of significant difference between lateralities or frequencies for the rise time tau_f. This result suggests that despite the larger MOCR Δ_{max} for the bilateral stimulation, the time course of the MOCR effect on the periphery is the same as unilateral stimulation, at least during the fast onset phase. The slow rise time constant, tau_s, however, was different between lateralties. This is largely driven by the higher tau_s registered in the right ears, the reasons for which are unclear. We randomized the probes between the right and left ears of participants, therefore this discrepancy is likely not measurement system related. It should be noted that the *tau* estimates reported here are partly dependent on the upper bound set to the two-term exponential fitting formula, 0.5 s for tau_f and 50 s for tau_s. These upper bounds were set based on prior physiological data (Liberman et al., 1996). For some fits, the *tau* was essentially at this bound, likely due to a prolonged evolution of the reflex over time. This occurred in ∼11% (participants × frequencies; 13 × 7) of the fits in bilateral (∼12% in the right ear and ∼13% in the left ear) stimulation despite these fits passing the HHG test. It is possible that such nuances affect tau_s differently from tau_f. Further data are necessary to clarify such details.

Alternatively, with a higher degree of MOC activation (higher level) and/or better time resolution (faster rate), it is possible that *tau* estimation may be less variable. It is also possible that despite the average SNR for CEOAEs used in the fitting process being 26 dB, precise *tau* calculation may require even higher SNR. It should, however, be noted that it is not the *tau* estimation that is important for clinical translation of this approach. In fact, the usefulness of *tau* for the clinic is currently unknown. Instead, it is the statistically significant characteristic reduction in Δ, approximated by a two-term exponential function, that is critical to determine if the change in the CEOAE level is likely physiologically driven. This time-course is the biggest advantage of the present approach over conventional methods as a direct link between the Δ and the MOCR can be established with greater certainty.

4.2.2. MOCR Steady-State

The MOCR Δ_{max} was estimated at the end of the 2-s activation window. In a majority of the participants, the Δ reached a steady-state earlier than 2 s. In a minority of the participants, it appeared to continue evolving, albeit gradually, even at the end of 2 s. If this method were to be translated to the clinic, the stimulus must be kept brief, 1 s or less. Comparisons between Δ_{max} estimated at 1, 1.25, 1.5, and 2 s suggest that a 1-second-long activation window along with a necessary silence period of 0.25–0.5 s, to allow MOCR to return to baseline, would be sufficient for MOCR estimation. This block duration would mean a test time of roughly 8 min. With further developments in signal processing, there is potential to reduce this test time further.

4.2.3. MOCR Recovery

The CEOAE Δ_{max} return to baseline at the end of 1 s in the recovery window captures the decaying portion of MOCR activity. The lack of MOCR activation in the baseline window suggests that the 55 dB peSPL/20 Hz clicks in the recovery window should likely only capture the decay of the MOCR, (the fall time constant, tau_f) and not activate any further MOCR activity. This recovery provides additional evidence that the clicks in the activation window did elicit the MOCR. However, the change in CEOAE level in the recovery window was not as robust as it was in the activation window. Unlike the rise time *tau*, we were unable to perform any statistics on the fall time *tau* due to the sparseness in the data. This is likely due to many reasons. (1) The poor time resolution in the recovery window relative to the activation window (100 vs. 32 ms). (2) Clicks evoking CEOAEs in the recovery window were much lower in level relative to the activation window (55 vs. 80 dB peSPL). (3) Lower click levels meant that the SNR was also lower relative to the activation window; 17.5 vs. 26 dB. These reasons likely rendered a larger proportion of the data in the recovery window unusable.

With these caveats in mind, The average fall time reported here (0.25 s; average across lateralties) is longer than that of the 0.16 s of Backus and Guinan (2006). This discrepancy could simply be due to the aforementioned caveats. Perhaps a higher level click and/or a faster click rate and/or longer averaging would capture this decay more robustly. However, higher click levels and faster click rates would also activate the MOCR and would not allow for the MOCR to decay. Therefore, one of the disadvantages of the proposed method is that the decay of the MOCR cannot be captured adequately in a clinically feasible time frame. However, it should be called to notice that the estimating the MOCR recovery for clinical purposes is not necessary. If the MOCR is activated, it will revert to its baseline activity given adequate time post stimulation. From the present study, Backus and Guinan (2006), and physiological data (Liberman, 1988; Liberman et al., 1996), it appears that 0.5 s should be sufficient for the MOCR to recover. This recovery time, along with the time to reach steady-state (1–1.5 s) is critical for designing future, more rapid, time-course-based MOCR tests. Conservatively speaking, a single block of activation and recovery (silent interval) can be achieved within 1.5–2 s.

4.3. MOCR Activity Was Measurable in All Participants

For any clinical test, it is vitally important that the test indeed measures the activity of the system it was designed to measure. In this vein, the fact that 100% of the participants had MOCR activation in at least one frequency (>90% participants across all seven frequencies) suggests that the approach and the parameters used in the present study is a feasible measure of MOCR function. More importantly, (1) these activations have passed a rigorous statistical test (HHG), and (2) display the characteristic time-course as reported by other human (Backus and Guinan, 2006) and animal studies (Liberman et al., 1996). As such, the certainty that these Δ over time are of MOCR in origin is higher than methods that reduce Δ to a single data point.

4.4. MEMR Influence

Inadvertent activation of the MEMR may negatively impact the confidence in MOCR estimation as both reflexes follow a similar time-course and have a similar impact on OAEs. Clinical tympanometry was used to indirectly determine the threshold of MEMR in most prior studies. More recent studies have consistently shown that this approach is not fail-safe (Guinan et al., 2003; Zhao and Dhar, 2009; Boothalingam and Goodman, 2021), as clinical tympanometers are relatively less sensitive and may underestimate MEMR thresholds by up to 20 dB (Feeney and Keefe, 1999; Feeney et al., 2003). There have been more recent efforts to detect MEMR presence in a more sensitive fashion using stimulus frequency emission group delay (Guinan et al., 2003; Zhao and Dhar, 2011), stimulus reflectance-based cut-offs (Abdala et al., 2014; Boothalingam and Purcell, 2015; Boothalingam et al., 2018), MEMR critical thresholds (Mertes, 2020) and using resampling techniques (Goodman et al., 2013; Mertes and Goodman, 2016; Lewis, 2017). Using the same time-course method used in this study, Boothalingam and Goodman (2021) showed that MEMR can be detected as stimulus level change with a high degree of certainty, based on its characteristic exponential growth. In addition, this approach is particularly useful for the present study because both the MEMR and the MOCR are elicited using the same stimulus.

A larger issue, however, is even if MEMR is detected, it cannot be ascertained if it will influence MOCR estimates. Our data (see **Figure 4**) suggests that even if MEMR is active it does not always necessitate influence on MOCR estimates. However, when the activation is large, in this case, >1 dB, there appears to be an influence on the MOCR estimate at 1 kHz. This is an important finding that may aid in the development of potential critical thresholds for MEMR influence on MOCR estimates. Critical thresholds can be useful in clinical settings, but they may not be universally valid. For instance, although Mertes (2020) established a statistical critical threshold for possible MEMR elicited using a 60 dB SPL noise elicitor, he indicated that critical thresholds can be influenced by a myriad of variables, e.g., choice of elicitor, elicitor level, OAE evoking stimulus, OAE evoking stimulus level, etc. Furthermore, critical thresholds suffer from the same issue, that it cannot be known if stimulus changes that breached critical threshold will affect MOCR estimation. One way around this problem is to run correlations between MEMR and MOCR estimates, as done in this study, to determine MEMR influence. This approach, however, is not feasible at an individual level. Therefore, each clinic should develop its own critical thresholds for their specific set of equipment and stimulus parameters.

The time-course method may offer a particular advantage over conventional methods in determining if MEMR activation influences MOCR. If the stimulus reflectance is reduced, the amount of stimulus energy reaching the cochlea is increased. As a result, the amount of MOCR activation will also be increased due to the increased stimulus energy, leading to a similarly larger CEOAE inhibition, Δ. In contrast, because the reflectance is relatively increased at higher frequencies (>∼1.5 kHz), the stimulus reaching the cochlea is reduced. The amount of stimulus energy activating the MOCR is thus decreased, producing smaller CEOAE Δ. Notwithstanding the complications related to the interaction between CEOAE level, as a result of stimulus level changes due to variable reflectance, and MOCR inhibition of CEOAEs (Hood et al., 1996; Lewis, 2019), the time-course method may still be useful. This is because, while the size of Δ cannot distinguish the presence from the absence of MEMR activation, the direction of Δ change over time can. That is, at least at the higher frequencies where the stimulus reflectance increases over time, if the CEOAE Δ is completely driven by the MEMR, a similar increase in CEOAE Δ over time can be expected. Therefore, if we observe a CEOAE inhibition at these higher frequencies despite the increasing stimulus reflectance, an argument can be made that even if the MEMR was activated, it is not the predominant factor driving the CEOAE Δ over time. With an appropriate MEMR detection method and a critical threshold in place, MOCR estimates can be considered with more confidence. Furthermore, the results from our data suggest that despite relatively large MEMR activation in n13 at all frequencies, the influence on MOCR appears to be present only at 1 kHz. As such, it is possible that MEMR effects on MOCR is minimal at

frequencies above 1.5 kHz. Further studies that use frequency-specific stimulation (e.g., tonebursts) may be able to shed further light on this conjecture.

5. CONCLUSION

We have introduced a time-course-based method of the MOCR magnitude (absolute Δ_{max}) and kinetics estimated solely using clicks without any additional elicitors. The following highlights from our findings suggest that our proposed method can be successful in clinical translation. (1) 100% of the participants had MOCR activation in at least one frequency among seven 1/3rd bands (>90% across all seven frequencies). (2) The mean MOCR Δ_{max} during the bilateral activation (1.69 dB) is commensurate with the 1–2 dB OAE inhibition typically reported across MOCR studies using contralateral noise. (3) MOCR kinetics are commensurate with prior reports using SF- and DPOAEs (Kim et al., 2001; Backus and Guinan, 2006). (4) The higher-than-typical click level is advantageous in generating high SNR (Lewis, 2019). (5) Use of statistical tests allow for objective detection of MOCR activity. (6) The ability to concurrently detect for MEMR contamination allows for greater confidence in our results. Future studies that compare the method proposed here with conventional OAE-based MOCR methods in a within-subjects design are required to directly establish the benefits of the proposed approach.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Institutional Review Board, Northwestern University. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

SB and SD designed the experiment. HM collected the data. SB, SD, and SG analyzed the data and wrote the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This research was supported by an American Speech-Language-Hearing Foundation New Investigators Research Grant and the Office of the Vice-Chancellor for Research and Graduate Education, University of Wisconsin-Madison to SB and a Knowles Center Grant to SD.

ACKNOWLEDGMENTS

Portions of this work were presented at the 2018 and 2019 American Auditory Society Scientific and Technical Conference, Scottsdale, AZ and the 2018 OAEvoke Workshop, Los Angeles, CA. We thank Chris Shera and Karolina Charaziak for help with bandlimited click generation. We thank Elizabeth Kunnel for help with data collection.

REFERENCES

Abdala, C., Dhar, S., Ahmadi, M., and Luo, P. (2014). Aging of the medial olivocochlear reflex and associations with speech perception. *J. Acoust. Soc. Am.* 135, 755–765. doi: 10.1121/1.4861841

Backus, B. C., and Guinan, J. J. (2006). Time-course of the human medial olivocochlear reflex. *J. Acoust. Soc. Am.* 119, 2889–2904. doi: 10.1121/1.2169918

Backus, B. C., and Guinan, J. J. (2007). Measurement of the distribution of medial olivocochlear acoustic reflex strengths across normal-hearing individuals via otoacoustic emissions. *J. Assoc. Res. Otolaryngol.* 8, 484–496. doi: 10.1007/s10162-007-0100-0

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate - a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B Methodol.* 57, 289–300. doi: 10.1111/j.2517-6161.1995.tb02031.x

Berlin, C. I., Hood, L. J., Hurley, A. E., Wen, H., and Kemp, D. T. (1995). Binaural noise suppresses linear click-evoked otoacoustic emissions more than ipsilateral or contralateral noise. *Hear. Res.* 87, 96–103. doi: 10.1016/0378-5955(95)00082-F

Berlin, C. I., Hood, L. J., Morlet, T., Wilensky, D., St John, P., Montgomery, E., et al. (2005). Absent or elevated middle ear muscle reflexes in the presence of normal otoacoustic emissions: a universal finding in 136 cases of auditory neuropathy/dys-synchrony. *J. Am. Acad. Audiol.* 16, 546–553. doi: 10.3766/jaaa.16.8.3

Bidelman, G. M., and Bhagat, S. P. (2015). Right-ear advantage drives the link between olivocochlear efferent "antimasking" and speech-in-noise listening benefits. *Neuroreport* 26, 483–487. doi: 10.1097/WNR.0000000000000376

Boero, L. E., Castagna, V. C., Di Guilmi, M. N., Goutman, J. D., Elgoyhen, A. B., and Gómez-Casati, M. E. (2018). Enhancement of the medial olivocochlear system prevents hidden hearing loss. *J. Neurosci.* 38, 7440–7451. doi: 10.1523/JNEUROSCI.0363-18.2018

Boothalingam, S., Allan, C., Allen, P., and Purcell, D. W. (2019). The medial olivocochlear reflex is unlikely to play a role in listening difficulties in children. *Trends Hear.* 23:2331216519870942. doi: 10.1177/2331216519870942

Boothalingam, S., and Goodman, S. S. (2021). Click evoked middle ear muscle Click evoked middle ear muscle reflex: spectral and temporal aspects. *J. Acoust. Soc. Am.* 149, 2628–2643. doi: 10.1121/10.0004217

Boothalingam, S., Kurke, J., and Dhar, S. (2018). Click-evoked auditory efferent activity: rate and level effects. *J. Assoc. Res. Otolaryngol.* 19, 421–434. doi: 10.1007/s10162-018-0664-x

Boothalingam, S., Macpherson, E., Allan, C., Allen, P., and Purcell, D. (2016). Localization-in-noise and binaural medial olivocochlear functioning in children and young adults. *J. Acoust. Soc. Am.* 139, 247–262. doi: 10.1121/1.4939708

Boothalingam, S., and Purcell, D. W. (2015). Influence of the stimulus presentation rate on medial olivocochlear system assays. *J. Acoust. Soc. Am.* 137, 724–732. doi: 10.1121/1.4906250

Borg, E., Nilsson, R., and Engström, B. (1983). Effect of the acoustic reflex on inner ear damage induced by industrial noise. *Acta Otolaryngol.* 96, 361–369. doi: 10.3109/00016488309132721

Brown, M. C. (1989). Morphology and response properties of single olivocochlear fibers in the guinea pig. *Hear. Res.* 40, 93–109. doi: 10.1016/0378-5955(89)90103-2

Charaziak, K. K., Dong, W., Altoé, A., and Shera, C. A. (2020). Asymmetry and microstructure of temporal-suppression patterns in basilar-membrane

responses to clicks: relation to tonal suppression and traveling-wave dispersion. *J. Assoc. Res. Otolaryngol.* 136(Pt 1):EL302–EL320. doi: 10.1007/s10162-020-00747-2

de Boer, J., and Thornton, A. R. D. (2007). Effect of subject task on contralateral suppression of click evoked otoacoustic emissions. *Hear. Res.* 233, 117–123. doi: 10.1016/j.heares.2007.08.002

Delano, P. H., Elgueda, D., Hamame, C. M., and Robles, L. (2007). Selective attention to visual stimuli reduces cochlear sensitivity in chinchillas. *J. Neurosci.* 27, 4146–4153. doi: 10.1523/JNEUROSCI.3702-06.2007

Feeney, M. P., and Keefe, D. H. (1999). Acoustic reflex detection using wide-band acoustic reflectance, admittance, and power measurements. *J. Speech Lang. Hear. Res.* 42, 1029–1041. doi: 10.1044/jslhr.4205.1029

Feeney, M. P., Keefe, D. H., Hunter, L. L., Fitzpatrick, D. F., Garinis, A. C., Putterman, D. B., et al. (2017). Normative wideband reflectance, equivalent admittance at the tympanic membrane, and acoustic stapedius reflex threshold in adults. *Ear Hear.* 38, e142–e160. doi: 10.1097/AUD.0000000000000399

Feeney, M. P., Keefe, D. H., and Marryott, L. P. (2003). Contralateral acoustic reflex thresholds for tonal activators using wideband energy reflectance and admittance. *J. Speech Lang. Hear. Res.* 46, 128–136. doi: 10.1044/1092-4388(2003/010)

Galambos, R., and Rupert, A. (1959). Action of the middle ear muscles in normal cats. *J. Acoust. Soc. Am.* 31, 349–355. doi: 10.1121/1.1907723

Gifford, M. L., and Guinan, J. J. (1987). Effects of electrical stimulation of medial olivocochlear neurons on ipsilateral and contralateral cochlear responses. *Hear. Res.* 29, 179–194. doi: 10.1016/0378-5955(87)90166-3

Glasberg, B. R., and Moore, B. C. J. (1990). Derivation of auditory filter shapes from notched-noise data. *Hear. Res.* 47, 103–138. doi: 10.1016/0378-5955(90)90170-T

Goodman, S. S. (2017). *Auditory Research Lab Audio Software*. https://github.com/myKungFu/ARLas (accessed May 1, 2019).

Goodman, S. S., Boothalingam, S., and Lichtenhan, J. T. (2021). Medial olivocochlear reflex effects on amplitude growth functions of long- and short-latency components of click-evoked otoacoustic emissions in humans. *J. Neurophysiol.* 125, 1938–1953. doi: 10.1152/jn.00410.2020

Goodman, S. S., Fitzpatrick, D. F., Ellison, J. C., Jesteadt, W., and Keefe, D. H. (2009). High-frequency click-evoked otoacoustic emissions and behavioral thresholds in humans. *J. Acoust. Soc. Am.* 125, 1014–1032. doi: 10.1121/1.3056566

Goodman, S. S., Mertes, I. B., Lewis, J. D., and Weissbeck, D. K. (2013). Medial olivocochlear-induced transient-evoked otoacoustic emission amplitude shifts in individual subjects. *J. Assoc. Res. Otolaryngol.* 14, 829–842. doi: 10.1007/s10162-013-0409-9

Guinan, J. J. (2006). Olivocochlear efferents: anatomy, physiology, function, and the measurement of efferent effects in humans. *Ear Hear.* 27, 589–607. doi: 10.1097/01.aud.0000240507.83072.e7

Guinan, J. J. (2014). Olivocochlear efferent function: issues regarding methods and the interpretation of results. *Front. Syst. Neurosci.* 8:142. doi: 10.3389/fnsys.2014.00142

Guinan, J. J., Backus, B. C., Lilaonitkul, W., and Aharonson, V. (2003). Medial olivocochlear efferent reflex in humans: otoacoustic emission (OAE) measurement issues and the advantages of stimulus frequency OAEs. *J. Assoc. Res. Otolaryngol.* 4, 521–540. doi: 10.1007/s10162-002-3037-3

Guinan, J. J., and Gifford, M. L. (1988). Effects of electrical stimulation of efferent olivocochlear neurons on cat auditory-nerve fibers. I. Rate-level functions. *Hear. Res.* 33, 97–113. doi: 10.1016/0378-5955(88)90023-8

Heller, R., Heller, Y., and Gorfine, M. (2013). A consistent multivariate test of association based on ranks of distances. *Biometrika* 100, 503–510. doi: 10.1093/biomet/ass070

Hood, L. J., Berlin, C. I., Bordelon, J., and Rose, K. (2003). Patients with auditory neuropathy/dys-synchrony lack efferent suppression of transient evoked otoacoustic emissions. *J. Am. Acad. Audiol.* 14, 302–313. doi: 10.1055/s-0040-1715746

Hood, L. J., Berlin, C. I., Hurley, A., Cecola, R. P., and Bell, B. (1996). Contralateral suppression of transient-evoked otoacoustic emissions in humans: intensity effects. *Hear. Res.* 101, 113–118. doi: 10.1016/S0378-5955(96)00138-4

James, A. L., Harrison, R. V., Pienkowski, M., Dajani, H. R., and Mount, R. J. (2005). Dynamics of real time DPOAE contralateral suppression in chinchillas and humans. *Int. J. Audiol.* 44, 118–129. doi: 10.1080/14992020400029996

Jerger, J., Burney, P., Mauldin, L., and Crump, B. (1974). Predicting hearing loss from the acoustic reflex. *J. Speech Hear. Disord.* 39, 11–22. doi: 10.1044/jshd.3901.11

Khalfa, S., Morlet, T., Micheyl, C., Morgon, A., and Collet, L. (1997). Evidence of peripheral hearing asymmetry in humans: clinical implications. *Acta Otolaryngol.* 117, 192–196. doi: 10.3109/00016489709117767

Killan, E. C., Brooke, R. E., Farrell, A., and Merrett, J. (2017). Clinically relevant long-term reliability of contralateral suppression of click-evoked otoacoustic emissions. *J. Hear. Sci.* 7, 27–36. doi: 10.17430/902926

Kim, D. O., Dorn, P. A., Neely, S. T., and Gorga, M. P. (2001). Adaptation of distortion product otoacoustic emission in humans. *J. Assoc. Res. Otolaryngol.* 2, 31–40. doi: 10.1007/s101620010066

Knudson, I. M., Shera, C. A., and Melcher, J. R. (2014). Increased contralateral suppression of otoacoustic emissions indicates a hyperresponsive medial olivocochlear system in humans with tinnitus and hyperacusis. *J. Neurophysiol.* 112, 3197–3208. doi: 10.1152/jn.00576.2014

Lauer, A. M., and May, B. J. (2011). The medial olivocochlear system attenuates the developmental impact of early noise exposure. *J. Assoc. Res. Otolaryngol.* 12, 329–343. doi: 10.1007/s10162-011-0262-7

Lewis, J. D. (2017). Synchronized spontaneous otoacoustic emissions provide a signal-to-noise ratio advantage in medial-olivocochlear reflex assays. *J. Assoc. Res. Otolaryngol.* 8, 1–13. doi: 10.1007/s10162-017-0645-5

Lewis, J. D. (2019). The effect of otoacoustic emission stimulus level on the strength and detectability of the medial olivocochlear reflex. *Ear Hear.* 40, 1391–1403. doi: 10.1097/AUD.0000000000000719

Liberman, M. C. (1988). Response properties of cochlear efferent neurons: monaural vs. binaural stimulation and the effects of noise. *J. Neurophysiol.* 60, 1779–1798. doi: 10.1152/jn.1988.60.5.1779

Liberman, M. C. (1990). Effects of chronic cochlear de-efferentation on auditory-nerve response. *Hear. Res.* 49, 209–223. doi: 10.1016/0378-5955(90)90105-X

Liberman, M. C., and Brown, M. C. (1986). Physiology and anatomy of single olivocochlear neurons in the cat. *Hear. Res.* 24, 17–36. doi: 10.1016/0378-5955(86)90003-1

Liberman, M. C., and Guinan, J. J. (1998). Feedback control of the auditory periphery: anti-masking effects of middle ear muscles vs. olivocochlear efferents. *J. Commun. Disord.* 31, 471–483. doi: 10.1016/S0021-9924(98)00019-7

Liberman, M. C., Liberman, L. D., and Maison, S. F. (2014). Efferent feedback slows cochlear aging. *J. Neurosci.* 34, 4599–4607. doi: 10.1523/JNEUROSCI.4923-13.2014

Liberman, M. C., Puria, S., and Guinan Jr, J. J. (1996). The ipsilaterally evoked olivocochlear reflex causes rapid adaptation of the 2 f1- f2 distortion product otoacoustic emission. *J. Acoust. Soc. Am.* 99, 3572–3584. doi: 10.1121/1.414956

Lilaonitkul, W., and Guinan, J. J. (2009a). Human medial olivocochlear reflex: effects as functions of contralateral, ipsilateral, and bilateral elicitor bandwidths. *J. Assoc. Res. Otolaryngol.* 10, 459–470. doi: 10.1007/s10162-009-0163-1

Lilaonitkul, W., and Guinan, J. J. (2009b). Reflex control of the human inner ear: a half-octave offset in medial efferent feedback that is consistent with an efferent role in the control of masking. *J. Neurophysiol.* 101, 1394–1406. doi: 10.1152/jn.90925.2008

Lilaonitkul, W., and Guinan, J. J. (2012). Frequency tuning of medial-olivocochlear-efferent acoustic reflexes in humans as functions of probe frequency. *J. Neurophysiol.* 107, 1598–1611. doi: 10.1152/jn.00549.2011

Lopez-Poveda, E. A. (2018). Olivocochlear efferents in animals and humans: from anatomy to clinical relevance. *Front. Neurol.* 9:276. doi: 10.3389/fneur.2018.00197

Mertes, I. B. (2020). Establishing critical differences in ear-canal stimulus amplitude for detecting middle ear muscle reflex activation during olivocochlear efferent measurements. *Int. J. Audiol.* 59, 140–147. doi: 10.1080/14992027.2019.1673491

Mertes, I. B., and Goodman, S. S. (2016). Within- and across-subject variability of repeated measurements of medial olivocochlear-induced changes in transient-evoked otoacoustic emissions. *Ear Hear.* 37, e72–e84. doi: 10.1097/AUD.0000000000000244

Mertes, I. B., Johnson, K. M., and Dinger, Z. A. (2019). Olivocochlear efferent contributions to speech-in-noise recognition across signal-to-noise ratios. *J. Acoust. Soc. Am.* 145, 1529–1540. doi: 10.1121/1.5094766

Mertes, I. B., and Leek, M. R. (2016). Concurrent measures of contralateral suppression of transient-evoked otoacoustic emissions and of auditory steady-state responsesa). *J. Acoust. Soc. Am.* 140, 2027–2038. doi: 10.1121/1.4962666

Mishra, S. K., and Lutman, M. E. (2013). Repeatability of click-evoked otoacoustic emission-based medial olivocochlear efferent assay. *Ear Hear.* 34, 789–798. doi: 10.1097/AUD.0b013e3182944c04

Morlet, T., Goforth, L., Hood, L. J., Ferber, C., Duclaux, R., and Berlin, C. I. (1999). Development of human cochlear active mechanism asymmetry: involvement of the medial olivocochlear system? *Hear. Res.* 137, 179–162. doi: 10.1016/S0378-5955(99)00155-0

Philibert, B., Veuillet, E., and Collet, L. (1998). Functional asymmetries of crossed and uncrossed medial olivocochlear efferent pathways in humans. *Neurosci. Lett.* 253, 99–102. doi: 10.1016/S0304-3940(98)00615-6

Rajan, R. (2000). Centrifugal pathways protect hearing sensitivity at the cochlea in noisy environments that exacerbate the damage induced by loud sound. *J. Neurosci.* 20, 6684–6693. doi: 10.1523/JNEUROSCI.20-17-06684.2000

Robertson, D., and Gummer, M. (1985). Physiological and morphological characterization of efferent neurones in the guinea pig cochlea. *Hear. Res.* 20, 63–77. doi: 10.1016/0378-5955(85)90059-0

Shera, C. A., Guinan, J. J., and Oxenham, A. J. (2002). Revised estimates of human cochlear tuning from otoacoustic and behavioral measurements. *Proc. Natl. Acad. Sci. U.S.A.* 99, 3318–3323. doi: 10.1073/pnas.032675099

Shera, C. A., Tubis, A., and Talmadge, C. L. (2008). Testing coherent reflection in chinchilla: auditory-nerve responses predict stimulus-frequency emissions. *J. Acoust. Soc. Am.* 124, 381–395. doi: 10.1121/1.2917805

Siegel, J. H., and Kim, D. O. (1982). Efferent neural control of cochlear mechanics? Olivocochlear bundle stimulation affects cochlear biomechanical nonlinearity. *Hear. Res.* 6, 171–182. doi: 10.1016/0378-5955(82)90052-1

Sridhar, T. S., Liberman, M. C., Brown, M. C., and Sewell, W. F. (1995). A novel cholinergic "slow effect" of efferent stimulation on cochlear potentials in the guinea pig. *J. Neurosci.* 15(5 Pt 1), 3667–3678. doi: 10.1523/JNEUROSCI.15-05-03667.1995

Srinivasan, S., Keil, A., Stratis, K., Woodruff Carr, K. L., and Smith, D. W. (2012). Effects of cross-modal selective attention on the sensory periphery: cochlear sensitivity is altered by selective attention. *Neuroscience* 223, 325–332. doi: 10.1016/j.neuroscience.2012.07.062

Stuart, A., and Cobb, K. M. (2015). Reliability of measures of transient evoked otoacoustic emissions with contralateral suppression. *J. Commun. Disord.* 58, 35–42. doi: 10.1016/j.jcomdis.2015.09.003

Valero, M. D., Hancock, K. E., Maison, S. F., and Liberman, M. C. (2018). Effects of cochlear synaptopathy on middle-ear muscle reflexes in unanesthetized mice. *Hear. Res.* 363, 109–118. doi: 10.1016/j.heares.2018.03.012

Veuillet, E., Collet, L., and Duclaux, R. (1991). Effect of contralateral acoustic stimulation on active cochlear micromechanical properties in human subjects: dependence on stimulus variables. *J. Neurophysiol.* 65, 724–735. doi: 10.1152/jn.1991.65.3.724

Walsh, E. J., McGee, J., McFadden, S. L., and Liberman, M. C. (1998). Long-term effects of sectioning the olivocochlear bundle in neonatal cats. *J. Neurosci.* 18, 3859–3869. doi: 10.1523/JNEUROSCI.18-10-03859.1998

Warr, W. B., and Guinan, J. J. (1979). Efferent innervation of the organ of corti: two separate systems. *Brain Res.* 173, 152–155. doi: 10.1016/0006-8993(79)91104-1

Warren, E. H., and Liberman, M. C. (1989). Effects of contralateral sound on auditory-nerve responses. II. Dependence on stimulus variables. *Hear. Res.* 37, 105–121. doi: 10.1016/0378-5955(89)90033-6

Wilson, U. S., Sadler, K. M., Hancock, K. E., Guinan, J. J., and Lichtenhan, J. T. (2017). Efferent inhibition strength is a physiological correlate of hyperacusis in children with autism spectrum disorder. *J. Neurophysiol.* 118, 1164–1172. doi: 10.1152/jn.00142.2017

Winslow, R. L., and Sachs, M. B. (1988). Single-tone intensity discrimination based on auditory-nerve rate responses in backgrounds of quiet, noise, and with stimulation of the crossed olivocochlear bundle. *Hear. Res.* 35, 165–189. doi: 10.1016/0378-5955(88)90116-5

Wojtczak, M., Beim, J. A., and Oxenham, A. J. (2017). Weak middle-ear-muscle reflex in humans with noise-induced tinnitus and normal hearing may reflect cochlear synaptopathy. *eNeuro* 4, 1–8. doi: 10.1523/ENEURO.0363-17.2017

Xing, D., and Gong, Q. (2017). Frequency specificity and left-ear advantage of medial olivocochlear efferent modulation: a study based on stimulus frequency otoacoustic emission. *Neuroreport* 28, 775–778. doi: 10.1097/WNR.0000000000000812

Zhao, W., and Dhar, S. (2009). The effect of contralateral acoustic stimulation on spontaneous otoacoustic emissions. *J. Assoc. Res. Otolaryngol.* 11, 53–67. doi: 10.1007/s10162-009-0189-4

Zhao, W., and Dhar, S. (2011). Fast and slow effects of medial olivocochlear efferent activity in humans. *PLoS ONE* 6:e18725. doi: 10.1371/journal.pone.0018725

Zhao, W., and Dhar, S. (2012). Frequency tuning of the contralateral medial olivocochlear reflex in humans. *J. Neurophysiol.* 108, 25–30. doi: 10.1152/jn.00051.2012

Zweig, G., and Shera, C. A. (1995). The origin of periodicity in the spectrum of evoked otoacoustic emissions. *J. Acoust. Soc. Am.* 98, 2018–2047. doi: 10.1121/1.413320

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Boothalingam, Goodman, MacCrae and Dhar. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 09 December 2021
doi: 10.3389/fnins.2021.759219

The Strength of the Medial Olivocochlear Reflex in Chinchillas Is Associated With Delayed Response Performance in a Visual Discrimination Task With Vocalizations as Distractors

*Sergio Vicencio-Jimenez[1,2,3], Giuliana Bucci-Mansilla[3], Macarena Bowen[3,4], Gonzalo Terreros[5], David Morales-Zepeda[3], Luis Robles[3] and Paul H. Délano[1,3,6,7]**

[1] Departamento de Otorrinolaringología, Hospital Clínico de la Universidad de Chile, Santiago, Chile, [2] Department of Otolaryngology-Head and Neck Surgery, The Center for Hearing and Balance, Johns Hopkins University School of Medicine, Baltimore, MD, United States, [3] Departamento de Neurociencia, Facultad de Medicina, Universidad de Chile, Santiago, Chile, [4] Departamento de Fonoaudiología, Facultad de Medicina, Universidad de Chile, Santiago, Chile, [5] Instituto de Ciencias de la Salud, Universidad de O'Higgins, Rancagua, Chile, [6] Facultad de Medicina, Biomedical Neuroscience Institute, Universidad de Chile, Santiago, Chile, [7] Centro Avanzado de Ingeniería Eléctrica y Electrónica, AC3E, Universidad Técnica Federico Santa María, Valparaíso, Chile

OPEN ACCESS

Edited by:
David K. Ryugo,
Garvan Institute of Medical Research, Australia

Reviewed by:
Guang-Wei Zhang,
University of Southern California, United States
Jochen Kaiser,
Goethe University Frankfurt, Germany

***Correspondence:**
Paul H. Délano
pdelano@med.uchile.cl

Specialty section:
This article was submitted to Auditory Cognitive Neuroscience, a section of the journal Frontiers in Neuroscience

Received: 16 August 2021
Accepted: 16 November 2021
Published: 09 December 2021

Citation:
Vicencio-Jimenez S, Bucci-Mansilla G, Bowen M, Terreros G, Morales-Zepeda D, Robles L and Délano PH (2021) The Strength of the Medial Olivocochlear Reflex in Chinchillas Is Associated With Delayed Response Performance in a Visual Discrimination Task With Vocalizations as Distractors. Front. Neurosci. 15:759219. doi: 10.3389/fnins.2021.759219

The ability to perceive the world is not merely a passive process but depends on sensorimotor loops and interactions that guide and actively bias our sensory systems. Understanding which and how cognitive processes participate in this active sensing is still an open question. In this context, the auditory system presents itself as an attractive model for this purpose as it features an efferent control network that projects from the cortex to subcortical nuclei and even to the sensory epithelium itself. This efferent system can regulate the cochlear amplifier sensitivity through medial olivocochlear (MOC) neurons located in the brainstem. The ability to suppress irrelevant sounds during selective attention to visual stimuli is one of the functions that have been attributed to this system. MOC neurons are also directly activated by sounds through a brainstem reflex circuit, a response linked to the ability to suppress auditory stimuli during visual attention. Human studies have suggested that MOC neurons are also recruited by other cognitive functions, such as working memory and predictability. The aim of this research was to explore whether cognitive processes related to delayed responses in a visual discrimination task were associated with MOC function. In this behavioral condition, chinchillas held their responses for more than 2.5 s after visual stimulus offset, with and without auditory distractors, and the accuracy of these responses was correlated with the magnitude of the MOC reflex. We found that the animals' performance decreased in presence of auditory distractors and that the results observed

in MOC reflex could predict this performance. The individual MOC strength correlated with behavioral performance during delayed responses with auditory distractors, but not without them. These results in chinchillas, suggest that MOC neurons are also recruited by other cognitive functions, such as working memory.

Keywords: delayed responses, working memory, otoacoustic emissions, chinchillas, olivocochlear, cognition

INTRODUCTION

Sensory perception is not just a passive phenomenon but involves the active participation of organism (Yang et al., 2016). In fact, in natural or ecological situations, the changes in the sensory organs are highly influenced by the internal process of the nervous system. This ranges from changes in our spatial relation with the environment to shifts in our focus or sensory priority. In this sense, our actions, cognition, and perception are interrelated, coupled in a sensorimotor cycle (Buhrmann et al., 2013; Di Paolo et al., 2017). Therefore, to understand the phenomenon of perception, it is necessary to know how our internal states and cognitive processes are associated with our sensory pathways. This is how top-down control pathways present themselves as compelling research targets for developing a better understanding of the cognitive control of perception.

In the auditory system, the efferent pathways form a neural network including the auditory cortex and subcortical nuclei, such as the thalamus, inferior colliculus, superior olivary complex (SOC), and cochlear nucleus (Malmierca and Ryugo, 2011; Elgueda and Delano, 2020). Through these efferent pathways, signals from the cerebral cortex can reach the cochlea *via* the olivocochlear (OC) system, which originates in the SOC (Rasmussen, 1946). In this context, it has been proposed that in cognitive processes like selective attention, descending signals modulate sensory responses at different levels of the nervous system (Johnson and Zatorre, 2006; Fritz et al., 2007; Lauer et al., 2021). For example, models of visual selective attention in the presence of auditory distractors have demonstrated changes in neural activity at different levels of the auditory pathway, including cortical regions (Woldorff et al., 1993; Shomstein and Yantis, 2004), subcortical nuclei (Hernández-Peón et al., 1956), the auditory nerve and the cochlear receptor (Delano et al., 2007). These changes in the cochlear and auditory afferent system functions have been attributed to modulations by the auditory corticofugal pathways (Aedo et al., 2016; Terreros et al., 2016). For instance, it has been shown that KO mice which lack efferent activity perform poorly on selective visual attention tasks in the presence of auditory distractors (Terreros et al., 2016). Furthermore, estimates of auditory efferent function in chinchillas [assessed by measuring the medial olivocochlear (MOC) reflex strength] have shown to predict visual attention performance in the presence of auditory distractors (Bowen et al., 2020). This evidence is also supported by findings in humans, which have reported modulations of otoacoustic emissions, a measure of cochlear hair cells activity, during visual selective attention (Wittekindt et al., 2014; Dragicevic et al., 2019).

All this information strongly supports the idea that the auditory descending pathways suppress irrelevant auditory stimuli when the organism focuses its attention on another sensory modality (such as vision). However, given the relevance of the olivocochlear system in the regulation of auditory input signals, it is not difficult to imagine that its cognitive control is not exclusively limited to sensory selective attention. Then, it is plausible that the OC system is sensitive to a wide variety of cognitive phenomena. For example, in a recent study, Marcenaro et al. (2021) showed evidence that the MOC reflex strength is modulated during visual working memory in humans. In this context, we investigated whether the MOC strength was associated with the behavioral performance of delayed responses (more than 2.5 s after stimulus offset) during a visual selective attention task in chinchillas. We compared the MOC reflex with the performance in quiet conditions and in the presence of two different types of auditory distractors: broadband noise (BBN) and chinchilla distress vocalizations. These were chosen for their difference in ecological relevance, with vocalizations being an ecologically more significant signal. Thus, we expected to find a correlation between performance in the presence of distractors and the MOC reflex, with vocalizations also having greater effects than BBN.

MATERIALS AND METHODS

Animals

We used a total of 19 adult male chinchillas (*Chinchilla laniger*, 4 ± 1 years of age) weighing between 500 and 700 g at the beginning of the behavioral training. Six animals were excluded from the analyses: three chinchillas dropped out of the training protocol due to health concerns, two others did not meet the behavioral criteria (see below), and in one it was not possible to perform adequate measurements of the MOC reflex. All chinchillas were housed in individual cages in a temperature and humidity-controlled room with a reverse light–dark cycle (lights on from 8 p.m. to 8 a.m.). In addition, they spent at least 3 h per week in an enrichment room, where they could exercise and bathe. They were given *ad libitum* access to water during the experimental period and were deprived of food, maintaining 85–90% of their previous *ad libitum* weight. All procedures were approved by the local Bioethics Committee (Animal Bioethics Committee, permit number 0844, Faculty of Medicine, University of Chile) and were performed according to the Care and Use of Laboratory Animals (National Research Council, 2011). These animals and raw data were used in our work published by Bowen et al. (2020). Here, we performed new analyses regarding late and the inter-trial time interval (ITI) responses to assess the cognitive processes associated with delayed behavioral responses.

Medial Olivocochlear Reflex Measurement

The strength of the MOC reflex was assessed by comparing values of distortion product otoacoustic emissions (DPOAEs), measured at 2f1–f2, in the absence and presence of broadband-noise contralateral acoustic stimulation (CAS; Maison and Liberman, 2000). DPOAEs were recorded using an ER-10B+ microphone system (Etymotic Research) with 40 dB gain, amplified 10,000×, filtered between 0.1 and 10 kHz (Krohn-Hite, model 3323), digitized with a 40 kHz sampling rate, and stored for off-line analysis.

Tests to estimate the MOC reflex were performed on awake animals on two separate occasions (test and retest, **Figures 1A,D**). Awake chinchillas were carefully placed in a soft body and neck restraint, keeping the room temperature at 23–24°C and with the lights off. Before performing the DPOAEs measurements, the animals underwent at least three habituation sessions to the body restrictor, in which the time in the restrictor was gradually increased. On average, the chinchillas endured this restriction for about 30–40 min, and movements were monitored with a video camera inside the acoustic chamber. In cases of excessive movement and discomfort, the session was aborted. The test and retest were measured in two different weeks. These tests comprised 1440 trials divided into three blocks of 480 trials: before, during, and after CAS. All experiments were controlled with custom programs developed in C language (LabWindows).

The auditory stimuli with which DPOAEs were elicited consisted of seven ipsilateral primary tone frequencies, delivered to the right ear, where f2 was equal to = 1440, 2040, 2884, 4080, 5769, 6125, and 8160 Hz. On the other hand, the contralateral BBN used for CAS had an intensity of ∼60 dB SPL and was delivered to the left ear. Both stimuli were digitally generated using two synchronized PCI cards (6.071-E, National Instruments) at 100,000 samples/s, attenuated with PA-5 programmable attenuators (System 3, Tucker-Davis Technologies) and delivered through ER-2 transducers (Etymotic Research) sealed to both external auditory meatus and pinna. Primary tones were presented at a frequency of 4 Hz with a duration of 15 ms, a rise/fall time of 5 ms, a fixed ratio of f2/f1 = 1.25 and L1/L2 = 65/60 dB SPL, with a delay of 200 ms. Contralateral non-continuous BBN (0.2–10 kHz) was administered at a presentation frequency of 4 Hz with a duration of 170 ms. At the beginning of each experiment, the sound pressure level in both ears was calibrated with an Etymotic® microphone.

Behavioral Apparatus and Training Procedures

The behavioral task was performed in an operant conditioning apparatus identical to the one used in Delano et al. (2007), located inside a double-wall room that attenuated sound. The training procedures were performed by experimenters who were blind to the MOC reflex values of the chinchillas. The time that was required for training (from the start to the entry into the experimental protocol) was approximately 2–3 months (**Figure 1B**). First, the animals had 1–2 weeks of habituation to the chamber and then began their training sessions. The chinchillas trained one session per day, 5 days per week. The task consisted of a two-choice visual discrimination test, which we have previously used in rats (Hamame et al., 2006), chinchillas (Delano et al., 2007; Bowen et al., 2020), and mice (Terreros et al., 2016; Jorratt et al., 2017). The operant conditioning chamber was located inside a double-wall room that attenuated external sound. The front panel of the apparatus had a central light (warning signal) located above the food dispenser and two sidelights (targets), each located above one of the response levers (right and left). A trial began with the onset of the central light (warning period) that lasted for 2 s, followed by the random onset of one of the target lights for a period of 0.5 s. Chinchillas were trained to respond by pressing the corresponding lever under the lateral light during the 5 s response period from the onset of the target light (**Figure 1C**). The ITI period varied randomly between 27 and 33 s. Correct responses during the response period were rewarded with a 45 mg pellet (Noyes PJNI-0045 Chinchilla Food Pellet; Research Diets, New Brunswick, NJ, United States). Early responses (pressing during the central light period), incorrect responses (pressing the opposite lever during the response period), and ITI responses (pressing after the response period was over) were punished with a 40-s time-out period, during which all lights were turned off. Trials in which chinchillas did not respond were defined as omissions and did not receive a time-out punishment. The behavioral variables measured were accuracy [correct responses/(correct responses + incorrect responses)], number of correct and incorrect responses, the number of omitted trials, and the latency of the lever press (time between the onset of the target light and the lever press). During the training period, the number of trials per session, the duration of the target light, the ITI period, and the punishment time were progressively modified according to the performance of the animals. After the chinchillas achieved an accuracy of at least 70% during a session of 110 trials with protocol values of 0.5 s target light duration, ITI of 27–33 s, and punishment of 40 s, they were recruited for the first day of the experimental protocol.

Experimental Protocol

The behavioral protocol consisted of 12 days of behavioral tests divided into three stages of 4 days with 110 trials each (Bowen et al., 2020; **Figure 1C**). On the first 4 days (baseline period), the chinchillas performed the two-choice discrimination task without auditory distractors, with the same parameters under which they finished the training stage. On days 5–8, the chinchillas performed the same visual discrimination task but in the presence of BBN as an auditory distractor (**Figure 1C**). Finally, on days 9–12, the chinchillas performed the visual discrimination task in the presence of an auditory distractor that consisted of alarm vocalizations from a male chinchilla (VOC) (**Figure 1C**).

Auditory Distractors

As mentioned above, we used two different auditory distractors during the experimental protocol: (1) a BBN (0.02–20 kHz) as an ecologically irrelevant distractor and (2) male chinchilla vocalizations as an ecologically relevant distractor. All

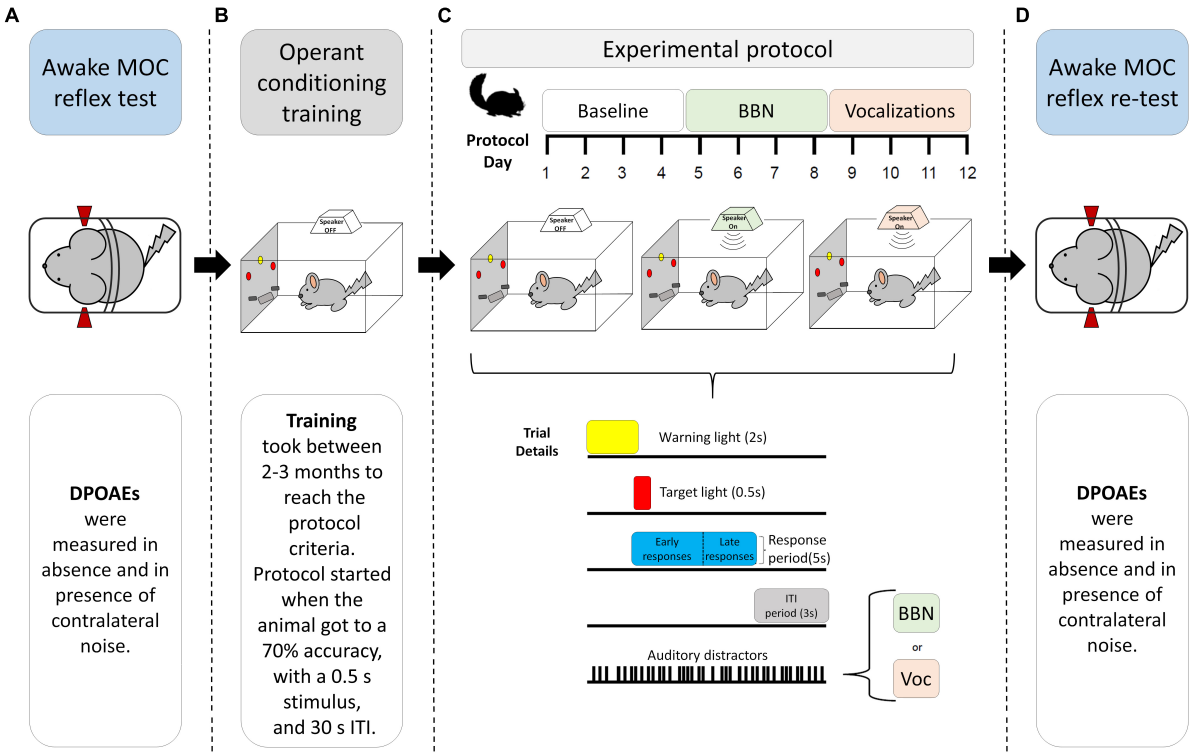

FIGURE 1 | Experimental design diagram. The image schematizes the experimental tests to which the animals were subjected. First, the MOC reflex was evaluated in awake chinchillas with restricted mobility **(A)**. After the evaluation, training of the visual attention task was started in the operant conditioning apparatus **(B)**. After 2–3 months of training, the animals moved to the 12-day experimental protocol **(C)**, where they performed the task in silence and in the presence of auditory distractors (BBN and Chinchilla vocalizations). The time course of the visual discrimination task can be seen at the bottom of panel **(C)**. The task began with a central warning light (2 s), followed by the presentation of a side light (0.5 s) indicating which response lever was the target of that trial. The sidelight also initiated the response time (5 s) for the animal to press the lever and receive a reward. The last 3 s shown in the diagram corresponded to part of the ITI-time response period, where the animal did not receive a reward if it pressed the lever. Finally, at the end of the 12-day protocol, the MOC reflex of the Chinchillas was measured again **(D)**.

vocalizations were previously recorded in a distress context and published by Moreno-Gómez et al. (2015). We used four clean harmonic male vocalizations (one for each of the 4 days with VOC) with the fundamental frequency (F0) between 538 and 861 Hz and dominant frequency around 1200 Hz. BBN and VOC distractors were presented binaurally at ∼65 dB SPL through a speaker (Sony, frequency response 20–20,000 Hz) located 1 m above the operant apparatus in free field conditions. Auditory distractors were delivered at an irregular rate centered at 2.5 ± 1.0 Hz (1.5–3.5 Hz, pseudo-randomly distributed) to prevent or diminish habituation.

Data Analysis

For the purposes of this study, we consider working memory as the transient representation of a signal during a period when the signal is no longer present and which serves to provide a subsequent response (Gazzaley and Nobre, 2012). Therefore, to study working memory, the response period was analyzed by looking at early and late time windows. The distinction between early and late time windows was determined by performing a frequency histogram for the latencies of the lever pressed. Using the temporal distribution of responses, the mean and standard deviation were calculated. With these parameters, we estimated the latency value that was more than two standard deviations away from the mean of the responses. That value was considered as the boundary between early and late responses. To estimate whether the average accuracy of animals during the basal period [the first stage of the behavioral protocol (**Figure 1C**)] was significantly greater than expected by chance, one-sample Student's *t*-test was performed against an expected mean of 0.5. A mixed effect analysis considered potential changes in the accuracy of animal responses on different days and conditions (Baseline, BBN, and VOC) during the experimental protocol. *Post hoc* analysis was performed using a Dunnett multiple comparisons test. The association between MOC reflex strength (CAS-induced DPOAE changes) and behavioral performance for each stage of the behavioral protocol was separately assessed using generalized linear models. The data was fitted using binomial family with a logit link was used. These procedures were like the ones used by Bowen et al. (2020). Data processing and statistical analyses were performed with MATLAB and GraphPad prism. Within the figures, the error bars correspond to the standard error of the mean (SEM). Statistical significance was defined as: $p > 0.05$ not significant (n.s.) and $p < 0.05$ as significant.

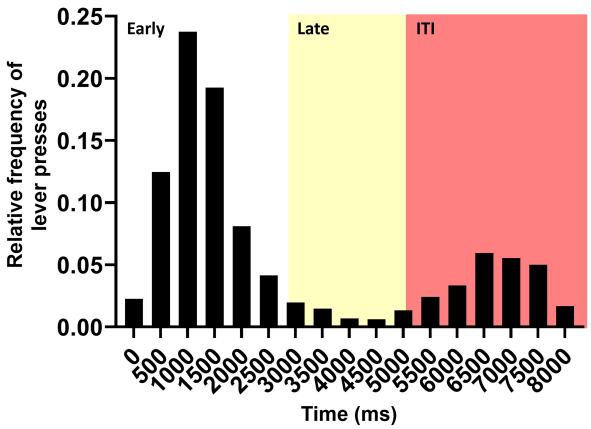

FIGURE 2 | Histogram showing the frequency of the responses during the baseline period of the protocol (days 1–4). The image shows the relative frequency of lever presses as a function of the trial elapsed time. Each bin of the histogram corresponds to a 500-ms window. The zero represents the beginning of the animals' response time (lateral light onset). The first 5 s represent the time in which the animal received a reward if it pressed the correct lever, which was separated into an early response window (white) and a late response window (yellow). The last 3 s correspond to the time in which the animal did not receive a reward for pressing the correct lever (red).

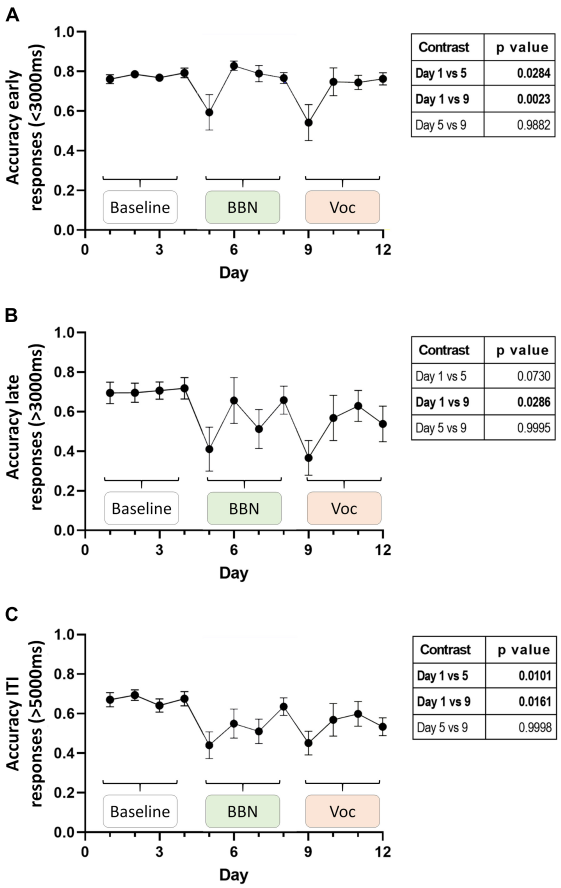

FIGURE 3 | Average accuracy of behavioral responses during the 12-day experimental protocol. Panel **(A)** shows the average accuracy of the responses from the early time window (0–3 s). In panel **(B)** the average accuracy of the responses of late time window (between 3 and 5 s), while in panel **(C)** the average accuracy of the ITI responses (between 5 and 8 s) is shown. Data are displayed as mean ± SEM (n = 13 chinchillas). Using linear mixed effect models, we found significant effects for behavioral performance on days 5 and 9 (see section "Results" in main text). A table with p-values from the Dunnett multiple comparisons test is shown to the right of each figure.

RESULTS

In this research, we evaluated the link between the MOC function and delayed responses in a visual discrimination task. For this, we observed the distribution of responses during the basal period of the protocol (**Figure 2**). We found that within the response period (0–5 s) the average lever press latency was 1.4 s with a standard deviation of 0.8 s. With this information, we defined the late responses as all those that were more than two standard deviations away from the mean latency. Thus, early responses corresponded to values between 0 and 3 s and late responses to values between 3 and 5 s from the lateral light onset (**Figure 2**, white and yellow, respectively). Late responses corresponded to approximately 5% of all lever presses. This time (at least 2.5 s post stimulus offset) is within the range of animal models of working memory (Wallace et al., 1980; Porritt and Poling, 2008; Lind et al., 2015). Along with this, when we looked at the totality of responses, we could see that ITI responses after the reward window were distributed between 5 and 8 s. No reward was received during ITI time, however, we found them interesting to analyze since they were made after receiving the lateral light stimulus, so the animal could potentially be making a late response to the signal. These responses could also involve working memory, considering that 4.5–7.5 s had elapsed after the stimulus offset. Thus, performance in these ITI responses (**Figure 2**, red) were also analyzed throughout the protocol.

Figure 3 shows the average performance of the animals during the 12-day protocol for each of these periods (**Figures 3A–C**, early, late, and ITI, respectively). As expected, during the basal period (days 1–4) of the early time window the chinchillas displayed high accuracy, averaging values close to 80% (0.78 ± 0.06). A mixed effect analysis found significant changes in accuracy throughout the experimental protocol [$F(11,132) = 3.954$, $p < 0.0001$], where a Dunnett's *post hoc* test found significant decreases for days 5 and 9 (**Figure 3A**, table insert). The decreases for these days, which corresponded to the first days of auditory distractors (BBN and vocalizations, respectively) are in line with those reported previously by Bowen et al. (2020). Inspection of the performance in the late time window also showed high accuracy in the responses during the basal period. In these 4 days, the animals averaged an accuracy of 0.7 ± 0.17. This is consistent with the hypothesis that the animals are using working memory during this period. In this time window, a mixed effect analysis also found significant changes in accuracy during the experimental protocol [$F(11,128) = 2.186$, $p = 0.0189$], where Dunnett's test found significant decreases only for day 9 of the protocol (**Figure 3B**, table insert). In

TABLE 1 | The generalized linear models evaluating the association between MOC reflex strength (DPOAE CAS-induced changes) and behavioral performance for early responses (<3 s).

Protocol day	Chi^2-statistic vs. constant model	DPOAE Frequency (Hz)	Estimate	SE	*t*-Stat	*p*-Value
Day 1 (Baseline-1)	5.03, *p*-value = 0.284	2884	−0.006233	0.030897	−0.20172	0.84013
		4080	0.057822	0.039265	1.4726	0.14085
		5768	0.03734	0.044147	0.84582	0.39765
		6125	−0.02995	0.052086	−0.57501	0.56529
Day 5 (BBN-1)	**32.8, *p*-value = 1.15e−05**	2884	0.01303	0.053587	0.24316	0.80788
		4080	**0.23955**	**0.048723**	**4.9165**	**8.812e−07**
		5768	−0.061187	0.059469	−1.0289	0.30353
		6125	−0.040424	0.067286	−0.60078	0.54799
Day 9 (VOC-1)	**18.4, *p*-value = 0.0101**	2884	0.053524	0.032191	1.6627	0.096375
		4080	**0.15633**	**0.045017**	**3.4728**	**0.0005151**
		5768	0.0068814	0.050159	0.13719	0.89088
		6125	**−0.1753**	**0.058562**	**−2.9934**	**0.0027593**

Significant results ($p < 0.05$ are bolded).

addition to the above, we found that the results in the ITI-time window followed a similar pattern to those found in the response window period (**Figure 3C**). On average, 67% of the animals' ITI responses during the basal period coincided with the lever signaled by the target light. In other words, if these responses had been within the reward window, the animals would have had 0.67 ± 0.1 accuracy. Moreover, these values differed significantly from the 0.5 accuracy expected only by chance [one sample *t*-test: $t = 7.334$, df = 12, $p < 0.0001$]. This evidence also implies that, even for these very late responses, chinchillas are using their memory to respond. Additionally, a mixed effect analysis found significant changes in accuracy during the experimental protocol [$F(11,131) = 2.993$, $p = 0.0014$], where Dunnett's *post hoc* test identified a significant decrease in responses associated with the correct lever on days 5 and 9 of the protocol (**Figure 3C**, table insert).

With these results, we investigated the association between performance in the visual discrimination task with auditory distractors and a measure of the MOC reflex. In the same way as Bowen et al. (2020), we used MOC reflex strength values obtained from awake chinchillas and evaluated the correlation of these values with the animal's performance at the three different periods during the behavioral protocol of 12 days. It is relevant to note that, in the absence of auditory distractors (during the baseline days), the MOC reflex was not a predictor of performance in any of the time windows (**Tables 1–3** and **Figure 4A**).

In the case of early responses (0–3 s) we found values very similar to those we previously reported (Bowen et al., 2020), identifying an association between the strength of the MOC reflex and the accuracy of the animals in the first days of the auditory distractor presentation (**Table 1**). For day 5 (BBN-1) the DPOAE amplitude at a frequency of 4080 Hz significantly correlated with task accuracy, while for day 9 (VOC-1) we found significant values for frequencies of 4080 and 6125 Hz (**Table 1**). On the other hand, analysis with generalized linear models for the late responses (3–5 s) found no significant link between the MOC reflex strength at different frequencies and the performance of the animals in the behavioral task (**Table 2**). In contrast, in the case of the ITI-time window (>5 s) we did find significant correlations between the strength of the MOC reflex and the accuracy of the animals on the first days of auditory distractor presentation (BBN-1 and VOC-1) (**Table 3**). We found that on day 5 (BBN-1) the DPOAE amplitude at a frequency of 4080 Hz correlated significantly with task accuracy, while for day 9 (VOC-1) the DPOAE at frequencies of 2884, 4080, and 6125 Hz were significantly associated with task accuracy (**Table 3**).

Figure 4 shows an example of the above mentioned correlations. It shows the individual values of accuracy in the three temporal windows as a function of the MOC reflex strength at 4080 Hz, for day 1 (Baseline-1, **Figure 4A**), day 5 (BBN-1, **Figure 4B**), and day 9 (VOC-1, **Figure 4C**). We chose the MOC reflex at this frequency because it was the one that presented the most significant associations with the behavioral results.

DISCUSSION

In the present work, we show that the functioning of MOC reflex, measured by the magnitude of DPOAEs suppression produced by contralateral noise in awake chinchillas, is associated with behavioral performance in a visual discrimination task with auditory distractors during periods in which working memory is relevant to accomplish the task. Specifically, we found that individual variability in the MOC reflex strength correlates with the accuracy of delayed responses (late and ITI responses, executed at more than 2.5 s from the target stimulus offset) in a visual discrimination task that was performed with chinchilla distress vocalizations as auditory distractors.

Our data were part of the same set used in a previous publication (Bowen et al., 2020), where we showed that the MOC reflex was a predictor of selective visual attention performance in the presence of auditory distractors. Here, we performed new analyses, including different periods in the behavioral task, which allowed us to study delayed responses. We divided correct responses into two periods: an early period (less than 2.5 s post

TABLE 2 | The generalized linear models evaluated the association between MOC reflex strength (DPOAE CAS-induced changes) and behavioral performance for late responses (3–5 s).

Protocol day	Chi^2-statistic vs. constant model	DPOAE Frequency (Hz)	Estimate	SE	*t*-Stat	*p*-Value
Day 1 (Baseline-1)	4.63, *p*-value = 0.327	2884	0.045465	0.131	0.34705	0.72855
		4080	−0.20551	0.15893	−1.293	0.196
		5768	−0.05293	0.20972	−0.25239	0.80074
		6125	0.30701	0.24406	1.258	0.20841
Day 5 (BBN-1)	5.04, *p*-value = 0.655	2884	0.24536	0.16866	1.4548	0.14574
		4080	−0.24856	0.2133	−1.1653	0.24389
		5768	−0.21825	0.30756	−0.70959	0.47796
		6125	0.48615	0.40453	1.2018	0.22945
Day 9 (VOC-1)	4.68, *p*-value = 0.699	2884	−0.00335	0.10564	−0.03175	0.97467
		4080	0.2029	0.09824	2.0653	0.038895
		5768	0.074641	0.13236	0.56393	0.5728
		6125	−0.18391	0.19457	−0.94521	0.34455

TABLE 3 | The generalized linear models evaluating the association between MOC reflex strength (DPOAE CAS-induced changes) and behavioral performance for ITI-time responses (>5 s).

Protocol day	Chi^2-statistic vs. constant model	DPOAE Frequency (Hz)	Estimate	SE	*t*-Stat	*p*-Value
Day 1 (Baseline-1)	3.64, *p*-value = 0.456	2884	−0.04567	0.048928	−0.93344	0.35059
		4080	0.017122	0.057903	0.2957	0.76746
		5768	0.055745	0.076765	0.72618	0.46773
		6125	−0.07975	0.09046	−0.88169	0.37795
Day 5 (BBN-1)	**18, *p*-value = 0.00124**	2884	−0.01456	0.062792	−0.23194	0.81658
		4080	**0.20648**	**0.080313**	**2.571**	**0.010142**
		5768	−0.13803	0.093076	−1.4829	0.13809
		6125	−0.12727	0.10404	−1.2233	0.2212
Day 9 (VOC-1)	**93, *p*-value = 3e−19**	**2884**	**0.74137**	**0.12628**	**5.8709**	**4.333e−09**
		4080	**0.5084**	**0.15067**	**3.3742**	**0.0007402**
		5768	0.14837	0.16142	0.91913	0.35803
		6125	**−1.0817**	**0.2384**	**−4.5373**	**5.699e−06**

Significant results ($p < 0.05$ are bolded).

target offset), and a late period (2.5–4.5 s post target offset). We also analyzed a third period of ITI responses, occurring 4.5–7.5 s after target offset, but with no reward and time-out punishment. The purpose of this separation was to have correct responses that were more related to the visual selective attention processes (early period), and responses that are probably related to visual and/or executive working memory (late correct and ITI). It is important to highlight, that in the cases of delayed responses (late correct and ITI), for having a good accuracy of discrimination, the visual target stimulus needed to be held for a few seconds in the memory buffer of the behaving chinchillas, indicating the use of visual, executive, or other type of working memory resource. As expected, the results in the early window were equivalent to those found by Bowen et al. (2020). Auditory distractors significantly decreased performance (**Figure 3A**) and MOC reflex values were good predictors for individual chinchilla performance (**Table 1** and **Figure 4**).

On the other hand, it is relevant to note that the results obtained in the late and ITI time window were like those we observed in the early window and, therefore, to those reported by Bowen et al. (2020). We observed that, in the basal response period of the experimental protocol, both in the late and ITI time window, the animals' correct responses were significantly greater than those expected by chance. Importantly, these results suggest that memory mechanisms are operating for late and ITI responses, allowing the animal to correctly press the lever associated with the brief target stimulus (0.5 s) that had disappeared more than 2.5 sec earlier. Furthermore, in these time windows we also found significant effects of auditory distractors on the animals' performance. For the late window, we only found a significant decrease for the first day of VOC distractor (day 9, **Figure 3B**). Potentially this is because, unlike BBNs, vocalizations are ecologically relevant signals, so they are expected to have a greater distracting effect. However, given the trend that can be observed in the data and how close it was to statistical significance, we believe that the failure to find significant changes on day 5 of the late window was probably due to the low number of trials we had within this window. For the case of the ITI-time window, we found significant decreases for both the first day of BBN (day 5) and the first VOC day (day 9, **Figure 3C**).

With these findings, we looked at whether the strength of the MOC reflex was associated with the performance of animals in the absence and presence of auditory distractors. These results are in line with what we have previously observed, for example,

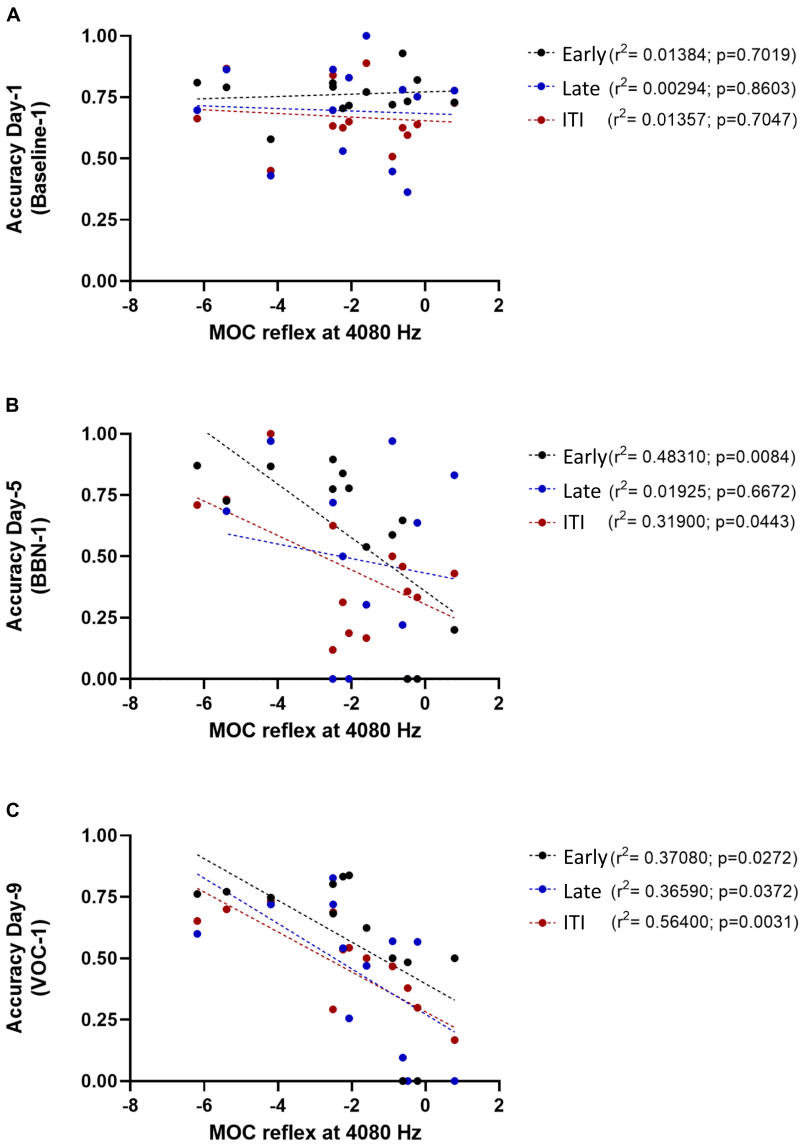

FIGURE 4 | Association between the MOC reflex (at 4080 Hz) and individual behavioral performance on different days of the behavioral protocol. The panel **(A)** represents the accuracy values on day 1 of the protocol (Baseline-1) as a function of the MOC reflex at 4080 Hz. In panel **(B)** it is shown the accuracy values on day 5 of the protocol (BBN-1) as a function of the MOC reflex at 4080, while panel **(C)** shows the values for day 9 of the protocol (VOC-1). The circles represent the individual values for each Chinchilla (n = 13) and the dotted lines correspond to the fitting curve. Black circles correspond to early responses, blue ones to the late response and red circles to ITI responses.

in mice where greater suppression of auditory nerve responses by contralateral noise was associated with better performance in a visual selective attention task with auditory distractors (Terreros et al., 2016).

In the early temporal window, we found that the strength of the MOC reflex (especially at 4080 Hz) predicted the performance of the animals in the presence of auditory distractors (**Table 1**). In contrast, generalized linear model analyses showed no significant correlation between individual MOC reflex values and animal performance in the late time window (3–5 s) (**Table 2**). Again, this probably relates with the low number of trials available for analysis. Especially because when we considered only the MOC 4080 Hz reflex value, a linear regression did yield significant effects at day 9 (VOC-1) (**Figure 4C**). This is supported by the fact that for the ITI-time period (with significantly more trials) the generalized linear model analysis found associations between the MOC reflex and the animals' responses in the presence of auditory distractors (**Table 3**). Moreover, these results were similar to those observed in the early period (**Figure 4**).

Despite the above, it is essential to highlight the limitations of this work. One of them is that the evaluation of the MOC reflex was performed at separate times from the experimental protocol. This implicitly assumes that the MOC reflex can be treated as a single, stable trait, which is not necessarily valid. This limits

us from establishing more direct relationships between both tests and leaves open the possibility of having obtained different MOC reflex values in the experimental conditions of the 12-day protocol. However, it is still noteworthy to have performed the measurement of the MOC reflex in fully awake animals, especially considering that evidence shows that strength of the MOC reflex is underestimated in anaesthetized chinchillas (Aedo et al., 2015). Also, and perhaps the most relevant limitation is that our results were obtained from a task that was not designed to assess working memory directly. We had to arbitrarily select late responses from a test initially designed for visual attention to a stimulus (Delano et al., 2007). Without going any further, almost 70% of the lever presses in the response window occurred in less than a second after the target light was turned off and only 5.3% occurred in the window that we defined as late. This explains the fact that our late window had such a low number of responses, and hence all the difficulties that this caused in the data analysis. Moreover, we cannot rule out that lever-related body orientation strategies may be biasing late responses, thus decreasing the working memory load. The same is true for the fact that we used the ITI responses as a proxy for valid late responses. Although it seems reasonable to assume that the animal was responding to the target stimulus (given the results observed in the basal period), the fact that these responses had no reward meant that their interpretation could not be completely equivalent to those that occurred within the valid response period. For the same reason, these experiments should be replicated in tests that are designed exclusively to evaluate WM, for example, by training animals to provide delayed responses with retractable levers to a stimulus. Finally, since we did not perform brain or cochlear function recordings during the task, nor any kind of functional manipulation, our ability to speculate about brain mechanisms is severely limited. For example, even assuming that our results for late responses are due to working memory, we cannot establish the potential neural networks involved (e.g., whether they are linked to visual processing areas, or perhaps to motor or premotor regions). Therefore, we believe that in the future it is necessary to study the brain dynamics associated with this type of behavior or to focus on manipulations that allow intervention of MOC function in behaving animals (e.g., with optogenetics or DREADD tools), in order to be able to dissect underlying mechanisms.

Concerning to visual working memory, our laboratory recently found evidence in humans showing that the MOC reflex is dynamically modulated when relevant visual stimuli are held in mind (Marcenaro et al., 2021). Together, with the present results in chinchillas, we propose a relationship between MOC function and working memory. In addition to our findings, Sörqvist et al. (2012) found modulation of the wave V of auditory brainstem responses during verbal–visual working memory, suggesting that top-down suppression of ascending auditory responses is important for the capacity of filtering distracting stimuli during working memory paradigms (Vogel et al., 2005; Sörqvist et al., 2012; Gaspar et al., 2016). Our results strengthen this notion, including the idea that the efferent function is also a predictor of performance in these tasks.

Therefore, our findings are in agreement with previous evidence that positions the olivocochlear efferent system as part of a dynamic network that is actively regulating sensory inputs as a function of the organism's relationship with the world and that is sensitive to cognitive states and experience (Oatman, 1971; Delano et al., 2007; Wittekindt et al., 2014; Terreros et al., 2016; Dragicevic et al., 2019; Bowen et al., 2020; Lauer et al., 2021; Marcenaro et al., 2021). This is also supported by anatomical evidence showing that, along with the auditory cortex, other cortical regions may interact directly with the efferent system (Olthof et al., 2019). Therefore, we believe it is pertinent to begin to expand the current understanding of the link between the efferent system and cognitive processes. In this context, the current evidence suggests, at least, this descending network is involved in sensory control associated with working memory, but that it is also possible to extend this notion to cognition as a global phenomenon. Moreover, in general terms, it makes sense to expect that any kind of relevant change in the individual–environment relationship will have consequences on the state of perceptual systems, on the signals entering the organism. Thus, we hypothesize that the auditory efferent control is probably related to the cognitive load of the organism, rather than to specific cognitive functions.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The animal study was reviewed and approved by the Animal Bioethics Committee, Faculty of Medicine, University of Chile.

AUTHOR CONTRIBUTIONS

MB, PD, SV-J, and GT: original idea and experimental conceptualization. MB, GB-M, and SV-J: experimental development. GB-M and SV-J: data analysis. SV-J: manuscript writing. GB-M, PD, DM-Z, LR, GT, and SV-J: manuscript editing. PD and GT: project administration and supervision. PD: funding acquisition. All authors contributed to the article and approved the submitted version.

FUNDING

ANID BASAL FB008, FONDEF ID20I10371, Proyecto Milenio ICN09_015, Vicerrectoría de Investigación y Desarrollo (VID) de la Universidad de Chile, código proyecto: ENL 19/20, and Fundación Guillermo Puelma.

ACKNOWLEDGMENTS

We thank Fernando Vergara and Pascal Jorratt for their technical assistance and Miguel Concha, for his valuable advice on data analysis. We also thank all the people who adopted chinchillas after their work in the laboratory was completed.

REFERENCES

Aedo, C., Tapia, E., Pavez, E., Elgueda, D., Delano, P. H., and Robles, L. (2015). Stronger efferent suppression of cochlear neural potentials by contralateral acoustic stimulation in awake than in anesthetized chinchilla. *Front. Syst. Neurosci.* 9:21. doi: 10.3389/fnsys.2015.00021

Aedo, C., Terreros, G., León, A., and Delano, P. H. (2016). The corticofugal effects of auditory cortex microstimulation on auditory nerve and superior olivary complex responses are mediated via alpha-9 nicotinic receptor subunit. *PLoS One* 11:e0155991. doi: 10.1371/journal.pone.0155991

Bowen, M., Terreros, G., Moreno-Gómez, F. N., Ipinza, M., Vicencio, S., Robles, L., et al. (2020). The olivocochlear reflex strength in awake chinchillas is relevant for behavioural performance during visual selective attention with auditory distractors. *Sci. Rep.* 10:14894. doi: 10.1038/s41598-020-71399-8

Buhrmann, T., Di Paolo, E., and Barandiaran, X. (2013). A dynamical systems account of sensorimotor contingencies. *Front. Psychol.* 4:285. doi: 10.3389/fpsyg.2013.00285

Delano, P. H., Elgueda, D., Hamame, C. M., and Robles, L. (2007). Selective attention to visual stimuli reduces cochlear sensitivity in chinchillas. *J. Neurosci.* 27, 4146–4153. doi: 10.1523/JNEUROSCI.3702-06.2007

Di Paolo, E., Buhrmann, T., and Barandiaran, X. (2017). *Sensorimotor Life: An Enactive Proposal*. Oxford: Oxford University Press, 304. doi: 10.1093/acprof:oso/9780198786849.001.0001

Dragicevic, C. D., Marcenaro, B., Navarrete, M., Robles, L., and Delano, P. H. (2019). Oscillatory infrasonic modulation of the cochlear amplifier by selective attention. *PLoS One* 14:e0208939. doi: 10.1371/journal.pone.0208939

Elgueda, D., and Delano, P. H. (2020). Corticofugal modulation of audition. *Curr. Opin. Physiol.* 18, 73–78. doi: 10.1016/j.cophys.2020.08.016

Fritz, J. B., Elhilali, M., David, S. V., and Shamma, S. A. (2007). Auditory attention - focusing the searchlight on sound. *Curr. Opin. Neurobiol.* 17, 437–455. doi: 10.1016/j.conb.2007.07.011

Gaspar, J. M., Christie, G. J., Prime, D. J., Jolicæur, P., and McDonald, J. J. (2016). Inability to suppress salient distractors predicts low visual working memory capacity. *Proc. Natl. Acad. Sci. U.S.A.* 113, 3693–3698. doi: 10.1073/pnas.1523471113

Gazzaley, A., and Nobre, A. C. (2012). Top-down modulation: bridging selective attention and working memory. *Trends Cogn. Sci.* 16, 129–135. doi: 10.1016/j.tics.2011.11.014

Hamame, C. M., Delano, P. H., and Robles, L. (2006). Relevance of a neutral cue in a two-choice detection task in the rat. *Biol. Res.* 39, 259–267. doi: 10.4067/S0716-97602006000200008

Hernández-Peón, R., Scherrer, H., and Jouvet, M. (1956). Modification of electric activity in cochlear nucleus during "Attention" in unanesthetized Cats. *Science* 123, 331–332. doi: 10.1126/science.123.3191.331

Johnson, J. A., and Zatorre, R. J. (2006). Neural substrates for dividing and focusing attention between simultaneous auditory and visual events. *Neuroimage* 31, 1673–1681. doi: 10.1016/j.neuroimage.2006.02.026

Jorratt, P., Delano, P. H., Delgado, C., Dagnino-Subiabre, A., and Terreros, G. (2017). Difference in perseverative errors during a visual attention task with auditory distractors in alpha-9 nicotinic receptor subunit wild type and knock-out mice. *Front. Cell. Neurosci.* 11:357. doi: 10.3389/fncel.2017.00357

Lauer, A. M., Jimenez, S. V., and Delano, P. H. (2021). Olivocochlear efferent effects on perception and behavior. *Hear. Res.* 2021:108207. doi: 10.1016/j.heares.2021.108207

Lind, J., Enquist, M., and Ghirlanda, S. (2015). Animal memory: a review of delayed matching-to-sample data. *Behav. Processes* 117, 52–58. doi: 10.1016/j.beproc.2014.11.019

Maison, S. F., and Liberman, M. C. (2000). Predicting vulnerability to acoustic injury with a noninvasive assay of olivocochlear reflex strength. *J. Neurosci.* 20, 4701–4707. doi: 10.1523/jneurosci.20-12-04701.2000

Malmierca, M., and Ryugo, D. (2011). "Descending connections of auditory cortex to the midbrain and brain stem," in *The Auditory Cortex*, eds A. Winer and C. E. Schreiner (New York, NY: Springer), 189–208.

Marcenaro, B., Leiva, A., Dragicevic, C., López, V., and Delano, P. H. (2021). The medial olivocochlear reflex strength is modulated during a visual working memory task. *J. Neurophysiol.* 125, 2309–2321. doi: 10.1152/jn.00032.2020

Moreno-Gómez, F., Leon, A., Velásquez, N., Penna, M., and Delano, P. (2015). Individual and sex distinctiveness in bark calls of domestic chinchillas elicited in a distress context. *J. Acoust. Soc. Am.* 138:1614. doi: 10.1121/1.4929750

National Research Council (2011). *Guide For The Care And Use Of Laboratory Animals*, 8th Edn. Washington, DC: The National Academies Press, doi: 10.17226/12910

Oatman, L. C. (1971). Role of visual attention on auditory evoked potentials in unanesthetized cats. *Exp. Neurol.* 32, 341–356. doi: 10.1016/0014-4886(71)90003-3

Olthof, B. M. J., Rees, A., and Gartside, S. E. (2019). Multiple nonauditory cortical regions innervate the auditory midbrain. *J. Neurosci.* 39, 8916–8928. doi: 10.1523/JNEUROSCI.1436-19.2019

Porritt, M., and Poling, A. (2008). Scopolamine effects under a titrating-delayed-nonmatching-to-position procedure. *Psychol. Rec.* 58, 37–49. doi: 10.1007/BF03395601

Rasmussen, G. L. (1946). The olivary peduncle and other fiber projections of the superior olivary complex. *J. Comp. Neurol.* 84, 141–219. doi: 10.1002/cne.900840204

Shomstein, S., and Yantis, S. (2004). Control of attention shifts between vision and audition in human cortex. *J. Neurosci.* 24, 10702–10706. doi: 10.1523/JNEUROSCI.2939-04.2004

Sörqvist, P., Stenfelt, S., and Rönnberg, J. (2012). Working memory capacity and visual–verbal cognitive load modulate auditory–sensory gating in the brainstem: toward a unified view of attention. *J. Cogn. Neurosci.* 24, 2147–2154. doi: 10.1162/jocn_a_00275

Terreros, G., Jorratt, P., Aedo, C., Elgoyhen, A. B., and Delano, P. H. (2016). Selective attention to visual stimuli using auditory distractors is altered in alpha-9 nicotinic receptor subunit knock-out mice. *J. Neurosci.* 36, 7198–7209. doi: 10.1523/JNEUROSCI.4031-15.2016

Vogel, E. K., McCollough, A. W., and Machizawa, M. G. (2005). Neural measures reveal individual differences in controlling access to working memory. *Nature* 438, 500–503. doi: 10.1038/nature04171

Wallace, J., Steinert, P. A., Scobie, S. R., and Spear, N. E. (1980). Stimulus modality and short-term memory in rats. *Anim. Learn. Behav.* 8, 10–16. doi: 10.3758/BF03209724

Wittekindt, A., Kaiser, J., and Abel, C. (2014). Attentional modulation of the inner ear: a combined otoacoustic emission and EEG study. *J. Neurosci.* 34, 9995–10002. doi: 10.1523/JNEUROSCI.4861-13.2014

Woldorff, M. G., Gallen, C. C., Hampson, S. A., Hillyard, S. A., Pantev, C., Sobel, D., et al. (1993). Modulation of early sensory processing in human auditory cortex during auditory selective attention. *Proc. Natl. Acad. Sci. U.S.A.* 90, 8722–8726. doi: 10.1073/pnas.90.18.8722

Yang, S. C.-H., Wolpert, D. M., and Lengyel, M. (2016). Theoretical perspectives on active sensing. *Curr. Opin. Behav. Sci.* 11, 100–108. doi: 10.1016/j.cobeha.2016.06.009

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Vicencio-Jimenez, Bucci-Mansilla, Bowen, Terreros, Morales-Zepeda, Robles and Délano. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 10 December 2021
doi: 10.3389/fnins.2021.717572

Listening in the Moment: How Bilingualism Interacts With Task Demands to Shape Active Listening

*Jennifer Krizman[1], Adam Tierney[2], Trent Nicol[1] and Nina Kraus[1,3]**

[1] Auditory Neuroscience Laboratory, Department of Communication Sciences and Disorders, Northwestern University, Evanston, IL, United States, [2] The ALPHALAB, Department of Psychological Sciences, Birkbeck, University of London, London, United Kingdom, [3] Departments of Neurobiology and Otolaryngology, Northwestern University, Evanston, IL, United States

While there is evidence for bilingual enhancements of inhibitory control and auditory processing, two processes that are fundamental to daily communication, it is not known how bilinguals utilize these cognitive and sensory enhancements during real-world listening. To test our hypothesis that bilinguals engage their enhanced cognitive and sensory processing in real-world listening situations, bilinguals and monolinguals performed a selective attention task involving competing talkers, a common demand of everyday listening, and then later passively listened to the same competing sentences. During the active and passive listening periods, evoked responses to the competing talkers were collected to understand how online auditory processing facilitates active listening and if this processing differs between bilinguals and monolinguals. Additionally, participants were tested on a separate measure of inhibitory control to see if inhibitory control abilities related with performance on the selective attention task. We found that although monolinguals and bilinguals performed similarly on the selective attention task, the groups differed in the neural and cognitive processes engaged to perform this task, compared to when they were passively listening to the talkers. Specifically, during active listening monolinguals had enhanced cortical phase consistency while bilinguals demonstrated enhanced subcortical phase consistency in the response to the pitch contours of the sentences, particularly during passive listening. Moreover, bilinguals' performance on the inhibitory control test related with performance on the selective attention test, a relationship that was not seen for monolinguals. These results are consistent with the hypothesis that bilinguals utilize inhibitory control and enhanced subcortical auditory processing in everyday listening situations to engage with sound in ways that are different than monolinguals.

Keywords: attention, listening, language, bilingualism, auditory

OPEN ACCESS

Edited by:
David Pérez-González, University of Salamanca, Spain

Reviewed by:
Linda Polka, McGill University, Canada
Jordi Costa-Faidella, University of Barcelona, Spain

***Correspondence:**
Nina Kraus
nkraus@northwestern.edu

Specialty section:
This article was submitted to Auditory Cognitive Neuroscience, a section of the journal Frontiers in Neuroscience

Received: *31 May 2021*
Accepted: *11 November 2021*
Published: *10 December 2021*

Citation:
Krizman J, Tierney A, Nicol T and Kraus N (2021) Listening in the Moment: How Bilingualism Interacts With Task Demands to Shape Active Listening. Front. Neurosci. 15:717572. doi: 10.3389/fnins.2021.717572

INTRODUCTION

Language experience leaves a pervasive imprint on the brain. Auditory-based language exposure not only supports language acquisition, but also facilitates the development of executive functions, namely inhibitory control (Wolfe and Bell, 2004; Figueras et al., 2008), working memory (Figueras et al., 2008; Conway et al., 2009; Gathercole and Baddeley, 2014), and sustained attention

(Mitchell and Quittner, 1996; Khan et al., 2005). Through their interconnected development, the executive and auditory systems become strongly tethered (Baddeley, 2003; Kral et al., 2016). This cognitive-sensory link is universal across spoken languages (Weissman et al., 2006; Wu et al., 2007), is supported by functional and structural connections between auditory and executive systems (Jürgens, 1983; Casseday et al., 2002; Raizada and Poldrack, 2007), and aids in focusing the attentional searchlight on a target sound (Fritz et al., 2007; Pichora-Fuller et al., 2016).

While language exposure facilitates development of auditory and executive systems in everyone (Conway et al., 2009; Kronenberger et al., 2020), the experience of learning two languages results in additional strengthening of the executive system in bilinguals (Mechelli et al., 2004; Abutalebi and Green, 2007; Abutalebi et al., 2011; Gold et al., 2013; Costa and Sebastián-Gallés, 2014). This strengthening is believed to result from the constant co-activation of both of their languages during communication (Spivey and Marian, 1999; Marian and Spivey, 2003; Thierry and Wu, 2007) and the resultant need to suppress the irrelevant language (Kroll et al., 2008; Van Heuven et al., 2008). Through the daily practice of selectively inhibiting one language, bilinguals fine-tune their inhibitory control ability (Bialystok and Viswanathan, 2009; Foy and Mann, 2014), an executive function that focuses attention on a relevant stimulus amid distractors. There is evidence that this daily tuning leads to bilingual advantages, relative to monolinguals, on tasks assessing inhibitory control [reviewed in Bialystok (2011), though some have failed to replicate this advantage, Paap et al., 2015; Dick et al., 2019].

In addition to aiding bilinguals in juggling their two languages, inhibitory control is important for all listeners during everyday communication. Everyday communication often takes place in noisy environments, requiring a listener to focus on a target talker amid distractors. When perceiving speech in noise, inhibitory control operates in concert with auditory processing to suppress irrelevant and enhance the representation of relevant stimulus features important for discriminating a target object from other sounds (Neill et al., 1995; Alain and Woods, 1999; Hopfinger et al., 2000; Tun et al., 2002; Wu et al., 2007). Because it contributes to pitch perception and aids in separating a target talker from distractors, the fundamental frequency (F0) is an important cue for perceiving speech in noise (Bregman et al., 1990; Darwin, 1997; F0; Bird and Darwin, 1998; Darwin et al., 2003). Indeed, more robust subcortical encoding of the F0, as measured by the frequency-following response (FFR), a neurophysiological response to sound generated predominately in the inferior colliculus (Chandrasekaran and Kraus, 2010; Coffey et al., 2016, 2019; Bidelman, 2018; White-Schwoch et al., 2019), relates with better speech-in-noise abilities (Anderson et al., 2010; Song et al., 2010). F0 encoding is malleable with language experience (Song et al., 2008; Krishnan et al., 2009). For example, bilinguals show enhanced subcortical encoding of the F0 that relates with their heightened inhibitory control ability (Krizman et al., 2012).

Given that bilinguals show enhancements in processes important for listening in the crowded acoustic environments commonly encountered in daily life, bilinguals would be expected to also show heightened speech-in-noise recognition. Previous literature, however, has found bilinguals struggle in this realm relative to monolinguals (Mayo et al., 1997; Cooke et al., 2008; Lecumberri et al., 2011; Lucks Mendel and Widner, 2016; Krizman et al., 2017; Morini, 2020). Despite bilinguals' cognitive and sensory enhancements, they perform more poorly than monolinguals on clinical assessments of listening to speech in noise (Shi, 2010, 2012; Stuart et al., 2010; Krizman et al., 2017; Skoe and Karayanidi, 2019). Interestingly, this perception-in-noise disadvantage only manifests when the target is linguistic; bilinguals instead show an advantage when the target is non-linguistic (i.e., a tone; Krizman et al., 2017). Given that bilinguals display impaired recognition in noise only for linguistic stimuli, and that bilinguals have enhanced inhibitory control (Bialystok, 2011) and F0 encoding (Krizman et al., 2012; Skoe et al., 2017), the bilingual speech-in-noise disadvantage may stem from difficulties with linguistic processing, which bilinguals may try to compensate for, at least partly, by strengthening the cognitive and sensory processes involved in these tasks (Crittenden and Duncan, 2014; Krizman et al., 2017; Skoe, 2019).

Given the evidence for enhanced inhibitory control and F0 encoding in bilinguals, these advantages may evince possible strategies for listening in noise that are uniquely successful for bilinguals. We hypothesize that the enhancements in inhibitory control and F0 encoding are the byproduct of continued reliance on these processes during everyday listening and reflect differences between monolinguals and bilinguals in how they understand speech, particularly degraded speech, such as speech in noise. To test whether bilinguals and monolinguals differ in the processes engaged to understand speech in noise, high-proficiency bilingual speakers of Spanish and English and monolingual speakers of English performed a selective attention task in which they were instructed to focus on one of two competing talkers, similar to the demands of everyday listening environments. We measured behavioral indices of task performance, and the neural processes engaged during the selective attention task were compared to neural processes engaged when the participants passively listened to these same competing sentences. Behaviorally, participants used a button box to select the correct button as instructed by the target talker amid competing instructions from the distracting talker. We predicted that bilinguals would perform more poorly than monolinguals on this task, consistent with bilinguals' poorer performance on speech-in-noise tests (Mayo et al., 1997; Shi, 2010, 2012).

Neurally, we used EEG to measure cortical and subcortical brain responses during the selective listening test and during passive exposure to the test sentences. We measured cortical neural entrainment across multiple frequency bands over the duration of the competing sentences and subcortical neural entrainment to the pitch contour of each talker. Cortically, active listening during a selective attention task increases neural entrainment relative to passive listening (Mesgarani and Chang, 2012; Golumbic et al., 2013; Ding and Simon, 2014). Evidence suggests that selective attention engages distinct cortical networks

in bilinguals and monolinguals (Olguin et al., 2019); however, it is unknown whether levels of cortical neural entrainment engaged during active or passive listening differs between these groups. Given the reported mechanistic differences (Astheimer et al., 2016; Olguin et al., 2019), we predicted that cortical neural entrainment during active listening would differ between bilinguals and monolinguals, while the groups would be matched during passive listening.

In contrast to cortical entrainment, work in animal models has shown that active listening decreases subcortical auditory encoding relative to passive listening (Slee and David, 2015). Despite these findings, the prevailing view in humans is that differences between active and passive listening are minimal or non-existent given early work showing a lack of attention effects on subcortical responses to simple auditory stimuli (e.g., clicks; Salamy and McKean, 1977; Collet and Duclaux, 1986) and the fact that, unlike cortical responses, subcortical responses can be reliably acquired whether the participant is awake or asleep (Osterhammel et al., 1985; Krishnan et al., 2005). Studies in humans have instead focused on whether differences can be seen in the subcortical response to the attended versus the ignored auditory stream during an active listening task, yielding mixed results (Galbraith et al., 1998; Varghese et al., 2015; Forte et al., 2017). As a first step to understanding the influence attention has on subcortical encoding during everyday listening situations and whether language experience impacts this influence, we wanted to focus on the general effect of attention on listening. Therefore, rather than comparing responses to the attended and ignored streams, we compared the groups when they were actively and passively listening to the talkers. Across all participants, we predicted that active listening would lead to a reduction in subcortical neural entrainment relative to passive listening, consistent with findings in animals (Slee and David, 2015). Moreover, given the previously reported bilingual enhancements in subcortical F0 encoding (Krizman et al., 2012; Skoe et al., 2017), we predicted that bilinguals would demonstrate greater subcortical neural entrainment to the pitch of the stimuli in both the active and passive listening conditions relative to monolinguals.

Separate from the selective-attention task, we tested participants on a measure of inhibitory control to determine whether inhibitory control abilities support performance on the selective attention task (Bialystok, 2015). We predicted bilinguals would outperform monolinguals on the inhibitory control measure, consistent with previous studies (Bialystok and Martin, 2004; Carlson and Meltzoff, 2008; Bialystok, 2009; de Abreu et al., 2012; Krizman et al., 2012, 2014). Additionally, because inhibitory control has been found to facilitate listening to a target talker during selective attention tasks (Alain and Woods, 1999; Tun et al., 2002), we predicted that performance on the inhibitory control and selective attention tasks would relate in both monolinguals and bilinguals. However, if bilinguals rely more heavily on inhibitory control during real-world listening (Krizman et al., 2012, 2017; Bialystok, 2015), then we expect this relationship between inhibitory control and active listening to be stronger in bilinguals.

MATERIALS AND METHODS

Participants

Participants were 40 adolescents and young adults [18.09 ± 0.64 years of age, 22 female, 19 low socioeconomic status (as indexed by maternal education, Hollingshead, 1975)], recruited from four Chicago high schools. The Northwestern University Institutional Review Board approved all procedures and consent was provided by participants 18 and older while informed written assent was given by adolescents younger than 18 and consent provided by their parent/guardian. Participants were monetarily compensated for their participation.

Participants were English monolinguals (n = 20; 55% female) and high-proficiency Spanish–English bilinguals (n = 20; 55% female) as measured by the Language Experience and Proficiency Questionnaire (LEAP-Q, Marian et al., 2007; Kaushanskaya et al., 2019). Maternal education level was used to approximate socioeconomic status (Hart and Risley, 1995; Hoff, 2003; D'Angiulli et al., 2008). Half of the monolinguals and 45% of the bilinguals had mothers with education levels of high-school graduate or below, while the remaining participants' maternal education levels were some college or beyond. To be included in this study, participants in both groups needed to have high English proficiency (≥7 out of 10 on English speaking and understanding proficiency, LEAP-Q). The monolinguals were required to have low Spanish proficiency (≤4 out of 10 on Spanish speaking and understanding proficiency, LEAP-Q), while the bilinguals were required to have high Spanish proficiency (≥6 out of 10 on Spanish speaking and understanding proficiency, LEAP-Q). Bilinguals were further required to have early acquisition of Spanish and English (≤5 years old). All subjects were required to have air conduction thresholds of <20 dB hearing level (HL) per octave for octaves from 125 to 8000 Hz and no diagnosis of a reading or language disorder. The two groups were matched on age [monolinguals: 18.07 ± 0.59 years, bilinguals: 18.12 ± 0.71 years; $F(1,38) = 0.049$, $p = 0.825$, $\eta_p^2 = 0.001$], sex (Kruskal–Wallis $X^2 = 0$, $p = 0.999$), maternal education level (Kruskal–Wallis $X^2 = 0.098$, $p = 0.755$), IQ [monolinguals: 104.65 ± 7.62; bilinguals: 101.65 ± 12.40; $F(1,38) = 0.850$, $p = 0.362$, $\eta_p^2 = 0.022$, Wechsler Abbreviated Scale of Intelligence, WASI, Wechsler, 1999], and English proficiency [$F(1,38) = 0.496$, $p = 0.486$, $\eta_p^2 = 0.013$], as determined from the LEAP-Q. As shown in **Table 1**, the groups differed on amount of daily English/Spanish exposure [$F(1,38) = 89.24$, $p < 0.0005$, $\eta_p^2 = 0.701$] and Spanish proficiency [$F(1,38) = 283.72$, $p < 0.0005$, $\eta_p^2 = 0.882$].

Inhibitory Control Task

Inhibitory control was assessed by the Integrated Visual and Auditory Continuous Performance Test (IVA + Plus[1], Richmond, VA, United States). This test is 20 min and is administered via a laptop computer. During this test 500 trials of 1's and 2's are visually or auditorily presented in a pseudo-random order. The participant clicks the mouse when a 1 (but not a 2) is seen or heard. Thus, the participant must attend to the

[1] www.braintrain.com

TABLE 1 | Language measures for English and Spanish in monolinguals and bilinguals.

		English	Spanish
English Monolingual	Age of acquisition	1.3 ± 1.34 years	n/a
	Proficiency	9.55 ±0.67	0.78 ± 1.41
	Exposure	95.25 ± 8.81	4.75 ± 8.81
Spanish–English Bilingual	Age of acquisition	2.40 ± 1.98 years	1.55 ± 1.80 years
	Proficiency	9.40 ± 0.68	8.08 ± 1.33
	Exposure	62.75 ± 12.62	37.25 ± 12.62

Proficiency is rated on a scale from 0 (none) to 10 (perfect) and reports of exposure within an individual sum to 100% across the two languages.

number while the modality is not a guiding cue to completing this task. Responses were converted to age-normed standard scores. These scores reflect how well the participant adapted to a change in modality when responding to 1's and ignoring 2's during the test. That is, a higher standard score reflects a smaller reaction time difference between modality switch and non-switch trials.

Competing-Talkers Selective Attention Task

Overview

Participants completed a selective attention task in which they listened to a target sentence presented simultaneously with a competing sentence. Modeled after the Coordinate Response Measure Corpus (Bolia et al., 2000), all sentences were of the format 'Ready [call sign] go to [color] [number] now.' Every trial consisted of two sentences, spoken simultaneously. One of the sentences was spoken by a female and one of the sentences was spoken by a male. One sentence had the call sign 'baron' and the other sentence had the call sign 'tiger.' The participant was assigned one of these call signs and was instructed to listen to the sentence that contained the target call sign. There was equal probability that the target call sign would be spoken by the male or female on any given trial. For the duration of a trial, four color-number combinations (e.g., red 3), arranged in the shape of an isosceles trapezoid, were projected on a screen in front of the participant. The four color-number options for each trial were (1) the target combination, (2) the competing color and the competing number, (3) the target color with the competing number, and (4) the competing color with the target number. At the end of the trial (i.e., after 'now,' during a 500 ms interstimulus interval) the participant selected a button that corresponded to the color-number combination that (s)he perceived using a hand-held response box with four buttons arranged in the same trapezoidal pattern. Evoked brain responses to the mixed sentences were collected to simultaneously measure online cortical and subcortical auditory processing during this selective attention task. Following the task, the participant's brain responses to the mixed sentences were recorded under a passive listening condition while the subject watched a muted cartoon.

Stimuli

To maximize differences in the spectral components of the competing sentences, sentences were constructed using natural utterances recorded at 44.1 kHz spoken by a female (average F0 = 220 Hz) and a male (average F0 = 137 Hz). For both the female and male sentences, a single exemplar of 'ready,' 'go,' 'to' and 'now' were used and 48 combinations of call sign (baron or tiger), color (red, blue, or green) and monosyllabic number (1, 2, 3, 4, 5, 6, 8, 9) were generated. Overlapping utterances (e.g., 'ready') were duration normalized in Audacity (Audacity 1.3.13[2]) between the two speakers and in the case of multiple possibilities (i.e., the call signs and the numbers) all potential utterances were duration normalized. This normalization ensured that all sentences would be of the same duration and that words would occur at the same time on every trial. To shorten collection time, utterances were compressed by 35% in Audacity (without altering the pitch). Words were root-mean-square normalized to 70 dB SPL, so that the signal-to-noise ratio was essentially 0 over the duration of each utterance. These individual utterances were then concatenated to form female and male versions of the 48 possible sentence combinations, each with a duration of 1970 ms (**Figure 1**). A sentence spoken by the female was then combined with a sentence spoken by the male and the mixed sentences were pseudo-randomly arranged for presentation with the caveats that no mixed sentence combination would be presented twice in succession, and that on any given trial the male and female were saying different colors and numbers. The same presentation order of these mixed sentences was used across all participants.

For analyses, the pitch contours of an average of the male sentences and, separately, an average of the female sentences were extracted in Praat[3] with the autocorrelation method using a silence threshold of 0.0003, a voicing threshold of 0.15, an octave cost of 0.01, an octave jump cost of 0.35, and a voiced/unvoiced cost of 0.04. These parameters were chosen to maximize the chances of identifying a continuous pitch contour for both voices.

Experimental Design

Setup

The participant sat in a comfortable chair in a sound-proof booth. Seven Ag/AgCl electrodes were affixed to the participant's scalp: active at central midline (Cz), frontal midline (Fz), and parietal midline (Pz), reference at the right and left earlobes, low forehead as ground, and a vertical eye channel placed below the left eye. Contact impedance was kept below 5 kΩ with interelectrode impedance differences < 3 kΩ. The participant was given a response box with four buttons that spatially matched the layout of the four possible color-number combinations projected on a screen in front of the participant on each trial. The mixed sentences were presented in alternating polarity at a rate of 0.4/s to both ears through insert earphones to the participant via NeuroSCAN's STIM2 presentation software (GenTask module, Compumedics, Inc.) at 70 dB SPL. Participants' behavioral and brain responses on each trial were recorded in SCAN 4.5 (Compumedics, Inc.) in continuous acquisition using an analog-to-digital conversion rate of 10 kHz and online filter settings of DC to 2000 Hz.

[2]http://audacity.sourceforge.net

[3]http://www.fon.hum.uva.nl/praat/

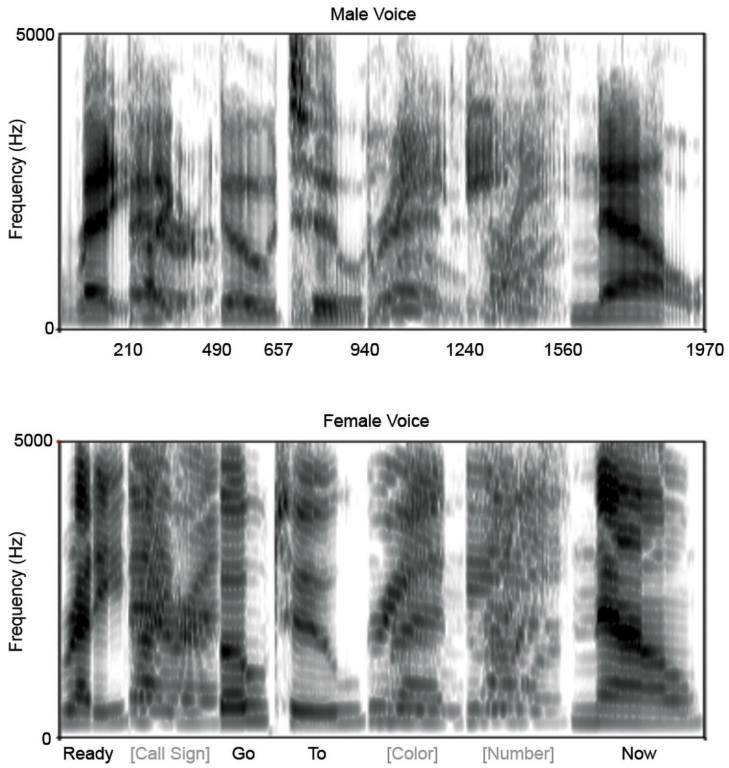

FIGURE 1 | Spectrograms of stimuli. For each voice, male **(top)** and female **(bottom)**, only one token of 'ready,' 'go,' 'to,' and 'now' were used, while two call signs, three colors, and multiple numbers were used. The segments of the spectrograms corresponding to the latter three are averages across the multiple utterances. The total utterance duration was 1,970 ms. Time labels on the *x*-axis of the male voice correspond to the word boundaries (e.g., 210 ms is when 'ready' ends and the call signs begin).

Training

After the electrodes were applied to the participant and instructions were given, testing began with training to familiarize the participant with the selective attention task. Training consisted of 12 blocks of 10 trials, where each trial is a mixed sentence (i.e., one sentence spoken by a male, giving instructions for one call sign, and one spoken by a female, giving instructions for the other call sign). For each block, the participant was instructed to attend to one call sign. At the end of the sentence, the participant pressed a button on the response box that corresponded to the color-number combination that (s)he was instructed to go to on that trial. Attended call sign alternated between blocks (e.g., 'Tiger' for block 1, 'Baron' for block 2, etc.). In any one trial, the attend call sign (e.g., 'Tiger') could be spoken by either the male or the female. Participants had to score 70% (i.e., 7/10) correct on two consecutive practice blocks or complete all 12 practice blocks to move on to the active task. Monolinguals and bilinguals did not differ on the number of lists required to achieve a passing score [monolinguals: 8.00 $\pm$ 3.74; bilinguals: 9.95 $\pm$ 3.39; $F(1,38) = 2.980$, $p = 0.092$, $\eta_p^2 = 0.073$].

Active Condition: Selective Attention to Competing Talkers

The active condition was identical to the training except that the active condition consisted of four blocks, with 500 trials comprising each block. During each block, the participant attended to one call sign ('Tiger' on blocks 1 and 3, 'Baron' on blocks 2 and 4). The male and female talkers each spoke the target sentence (i.e., the sentence containing the target call sign) 50% of the time during a block, resulting in 1000 trials where the male was attended and 1000 in which the female was attended across the entire active listening condition. Including breaks, this condition lasted about 90 min.

Passive Condition: Cartoon Watching While Hearing Competing Talkers

After the active task was completed, participant brain responses were collected during a passive listening condition. During this condition, participants were told that they no longer needed to pay attention to the competing talkers. Instead, they were instructed to watch a muted cartoon ('Road Runner and Friends' from *Looney Tunes Golden Collection Volume II*). Passive responses were recorded to 600 mixed sentences. These 600 mixed sentences were the same as the initial 600 sentences presented during the active listening condition. This condition lasted about 30 min.

Data Reduction and Processing

Offline, the triggers on both the active and passive continuous files were re-coded to reflect participant performance on the

active task (i.e., correct male attend, correct female attend, incorrect male attend, and incorrect female attend). Analyses were run on correctly attended trials to ensure that the analyzed trials were ones in which the participant was actively engaging with the stimuli during the selective attention task. To include all participants, only the first 175 correct trials were used (i.e., the lowest number available after factoring in task performance and artifact rejection of the subcortical responses as described below). Intertrial phase consistency of the brain response was analyzed using two different sets of parameters that reflected primarily low-frequency activity from the cortex and separately high-frequency activity from the auditory midbrain.

Cortical

To assess cortical neural entrainment, intertrial phase consistency over discrete frequency bands was calculated for active versus passive listening. First, data were downsampled to 500 Hz and spatial filtering was performed using singular value decomposition in Neuroscan Edit v4.3 to remove eyeblinks. Next, phase consistency calculations were performed over consecutive 200 ms sliding response windows with a 100 ms on and off hanning ramp (199 ms overlap) over the duration of the response, which provided us with 1-ms resolution of phase consistency. In each 200 ms window, a fast Fourier transform was used to calculate the spectrum of each trial. This calculation resulted in a vector for each frequency that contained a vector length, a reflection of encoding strength for each frequency, and a phase, which contained information about the timing of the response to that frequency. To examine the phase consistency of the response, each vector was transformed into a unit vector (i.e., amplitude information was removed) and then the first 175 vectors (i.e., trials) at each frequency were averaged so that the length of the resulting vector provided a measure of the intertrial phase consistency. Active versus passive comparisons were done on a composite of the attend male correct and attend female correct responses. Phase consistency values were then computed for theta (3–7 Hz), alpha (8–12 Hz), beta (20–30 Hz), gamma (31–50 Hz), and high gamma (65–150 Hz) frequency bands during the individual words in the sentence. The amplitude envelopes of the sentences have the highest energy at 4 Hz, which is within the theta band.

Subcortical

To assess subcortical neural entrainment, intertrial phase consistency to the pitch contour of the male voice and the pitch contour of the female voice were calculated on a composite of the attend male correct and attend female correct responses. These phase-consistency calculations were performed over a consecutive 40 ms sliding response window with 20 ms on and off hanning ramp (39 ms overlap) for the duration of the response, providing us with 1-ms resolution of phase consistency. Only responses that fell below the artifact rejection criterion ($\pm$50 μV) were included in the analyses. Phase-consistency calculations for subcortical frequencies were identical to calculation procedures for cortical frequencies. Phase-consistency values were computed for the frequencies ($\pm$2 Hz) comprising the pitch contour of the male voice, and separately the female voice. Phase consistency for each voice were calculated, using a 10-ms lag, for words that were consistent on every trial ('ready,' 'go,' 'to,' and 'now'). Using a 10 ms lag ensures that we are picking up on subcortical encoding of the pitch contour, given that this lag corresponds to the lag between stimulus and subcortical response that has been reported previously (Chandrasekaran and Kraus, 2010; Coffey et al., 2016). Parts of the sentence that contained multiple word options (e.g., number) were not analyzed because there was no consistent pitch contour to track across the multiple words (**Figure 1**).

Data Analyses

Behavioral Performance

Bilingual and monolingual groups were compared on their performance on the selective attention task and the inhibitory control test using a separate univariate analysis of variance (ANOVA) for each test. Additionally, performance on these behavioral tests was correlated within each language group to determine whether the participants relied upon their inhibitory control abilities to perform the selective attention task. Mean $\pm$1 standard deviation are reported for the two language groups on each measure.

Cortical Phase Consistency

Cortical responses were compared using a 2 (language group: Monolingual, Bilingual) × 2 (listening condition: Active, Passive) × 3 (electrode: Fz, Cz, Pz) × 7 (word: 'ready,' [call-sign], 'go,' 'to,' [color], [number], 'now') × 5 (frequency band: theta, alpha, beta, gamma, and high gamma) repeated measures analysis of variance (RMANOVA) to determine if cortical phase consistency differed between language groups for the active and passive listening conditions. This RMANOVA was followed up with a 2 (language group: Monolingual, Bilingual) × 2 (listening condition: Active, Passive) × 3 (Electrode: Fz, Cz, Pz) × 7 (Word: 'ready', [call-sign], 'go,' 'to,' [color], [number], 'now') RMANOVA for each frequency band. Significant interactions were further analyzed to characterize the effects. *Post hoc* analyses were corrected for multiple comparisons. Mean ± 1 standard deviation for the various measures are reported within the text in parentheses. Remaining tests are reported in the **Supplementary Material**.

Subcortical Phase Consistency

To identify differences in subcortical phase consistency as a function of language experience, subcortical phase consistency to the F0 was compared between active and passive listening conditions using a 2 (language group: Monolingual, Bilingual) × 2 (condition: Active, Passive) × 3 (electrode: Fz, Cz, Pz) × 2 (pitch contour: Male, Female) × 4 (word: 'ready,' 'go,' 'to,' 'now') RMANOVA. Significant interactions were analyzed further to characterize the effects. *Post hoc* analyses were corrected for multiple comparisons. Mean ± 1 standard deviation for the various measures are reported within the text in parentheses. Remaining tests are reported in the **Supplementary Material**.

Cortical – Subcortical Comparisons

Effects of Language Experience

To further investigate whether bilinguals and monolinguals engage different mechanisms for active and passive listening, we explored whether cortical and subcortical interactions during active and passive listening differed for the two groups. To do this, we averaged all active, and separately passive, cortical phase consistency data, collapsing across electrode, frequency band, and word for monolinguals and bilinguals. Subcortical active and passive phase consistency data was similarly averaged over electrode, pitch contour (i.e., talker), and word. To facilitate comparison between cortical and subcortical phase consistency, only the four words with consistent pitch contours across trials ('ready,' 'go,' 'to,' 'now') were included in these calculations. These composite values were then analyzed using a 2 (language group: Monolingual, Bilingual) × 2 (auditory region: Cortical, Subcortical) RMANOVA to determine whether there were differences in the way monolinguals and bilinguals utilized cortical and subcortical auditory processing when listening under different conditions.

Disentangling High Gamma and the Male Pitch Contour

High gamma (65 – 150 Hz) and phase consistency to the male pitch contour (average 137 Hz) overlap in frequency but are presumed to originate from cortical and subcortical sources, respectively (Edwards et al., 2005; Chandrasekaran and Kraus, 2010; Mesgarani and Chang, 2012; Bidelman, 2015; White-Schwoch et al., 2019). To determine whether we were capturing two distinct sources of activity, these data were analyzed using a 2 (response: high gamma, male pitch contour) × 2 (listening condition: Active, Passive) × 3 (electrode: Fz, Cz, Pz), by 4 (word: 'ready,' 'go,' 'to,' 'now') RMANOVA. To illustrate differences between cortical and subcortical consistency, consistency of the male pitch contour, high gamma, and the female pitch contour were plotted by word and listening condition, to visually depict the similarities and differences of the male pitch contour with respect to high gamma activity, which has known cortical generators (Edwards et al., 2005; Cervenka et al., 2011; Mesgarani and Chang, 2012), and the female pitch contour, whose frequency of 220 Hz is beyond cortical phase-locking abilities and thus reflects predominantly midbrain sources (Liang-Fa et al., 2006; Coffey et al., 2016, 2019; White-Schwoch et al., 2019).

RESULTS

In summary, monolinguals and bilinguals perform equivalently on the selective attention and inhibitory control tasks, but a relationship between performance on the two tasks exists only in bilinguals. Also, cortical consistency is enhanced in monolinguals, relative to bilinguals, especially during active listening. In contrast, subcortical consistency is enhanced in bilinguals relative to monolinguals, but is reduced during active relative to passive listening.

Behavioral

Monolingual and bilingual participants performed equivalently on the selective attention task [**Figure 2**, monolinguals: 53.80 ± 13.47%; bilinguals: 51.01 ± 12.13%; $F(1,38) = 0.47$, $p = 0.497$, $\eta_p^2 = 0.012$] as well as the inhibitory control test [monolinguals: 72.25 ± 38.13; bilinguals: 78.25 ± 42.00; $F(1,38) = 0.224$, $p = 0.639$, $\eta_p^2 = 0.006$]. While performance on these tests did not relate in monolinguals [$r(18) = -0.188$, $p = 0.427$], performance was related for bilinguals [$r(18) = 0.530$, $p = 0.016$], with better inhibitory control corresponding to

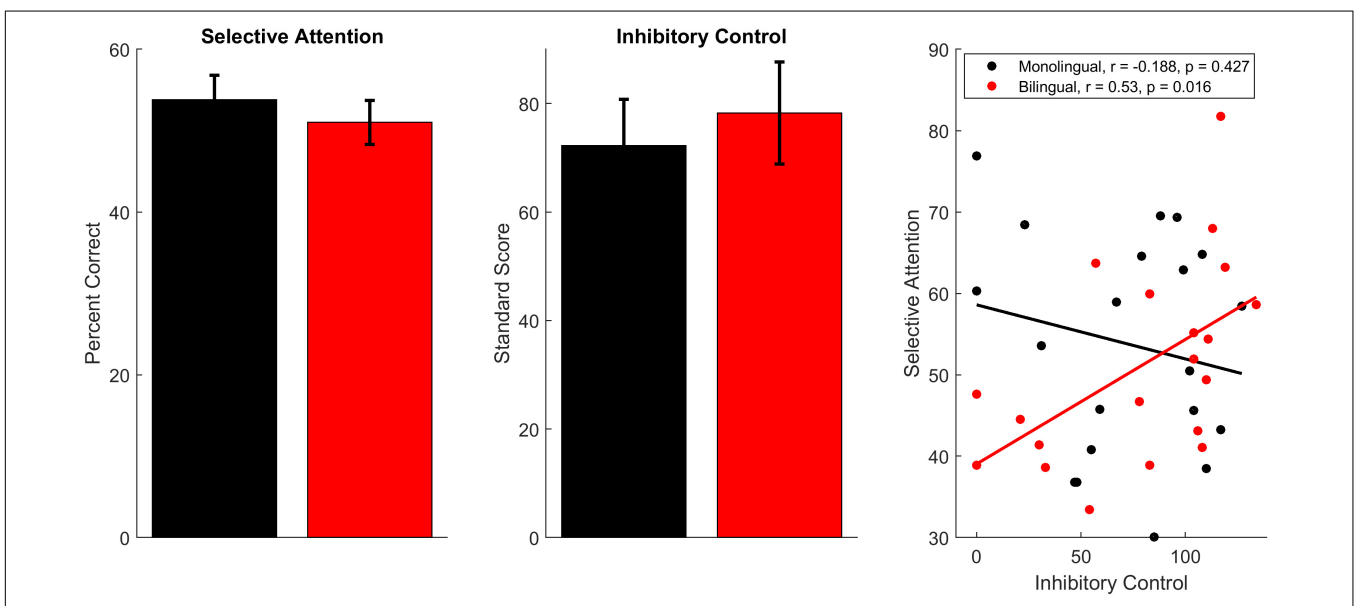

FIGURE 2 | Behavioral performance on the selective attention and inhibitory control tests. Bilinguals (red) and monolinguals (black) performed equivalently on both of these measures. However, performance on these measures was only correlated in bilingual participants. Error bars display standard error of the mean.

TABLE 2 | Cortical main effects and interactions for 2 (language group: monolingual, bilingual) × 2 (listening condition: active, passive) × 3 (electrode: Fz, Cz, Pz) × 7 (word: 'ready,' [call-sign], 'go,' 'to,' [color], [number], 'now') × 5[frequency band: theta (3–7 Hz), alpha (8–12 Hz), beta (20–30 Hz), gamma (31–50 Hz), and high gamma (65–150 Hz)] RMANOVA.

	F	df	*p*	η_p^2
Listening condition	**33.26**	**(1, 38)**	**<0.0005**	**0.467**
Language group	**4.65**	**(1, 38)**	**0.038**	**0.109**
Electrode	**55.82**	**(2, 76)**	**<0.0005**	**0.595**
Word	**42.63**	**(6, 228)**	**<0.0005**	**0.529**
Frequency band	**206.12**	**(4, 152)**	**<0.0005**	**0.844**
Listening condition × Language group	1.48	(1, 38)	0.232	0.037
Listening condition × Electrode	*2.55*	*(2, 76)*	*0.085*	*0.063*
Listening condition × Word	**5.69**	**(6, 228)**	**<0.0005**	**0.130**
Listening condition × Frequency band	**31.12**	**(4, 152)**	**<0.0005**	**0.450**
Electrode × Language group	1.73	(2, 76)	0.183	0.044
Electrode × Word	**5.42**	**(12, 456)**	**<0.0005**	**0.125**
Electrode × Frequency band	**29.40**	**(8, 304)**	**<0.0005**	**0.436**
Word × Language group	1.10	(6, 228)	0.366	0.028
Word × Frequency Band	**21.78**	**(24, 912)**	**<0.0005**	**0.364**
Frequency band × Language group	**5.21**	**(4, 152)**	**0.001**	**0.121**
Listening condition × Electrode × Language group	2.05	(2, 76)	0.135	0.051
Listening condition × Word × Language group	0.70	(6, 228)	0.646	0.018
Listening condition × Frequency band × Language group	1.83	(4, 152)	0.126	0.046
Listening condition × Electrode × Word	1.04	(12, 456)	0.412	0.027
Listening condition × Electrode × Frequency band	**2.23**	**(8, 304)**	**0.025**	**0.056**
Listening condition × Word × Frequency band	**6.72**	**(24, 912)**	**<0.0005**	**0.150**
Electrode × Word × Language group	1.01	(12, 456)	0.443	0.026
Electrode × Frequency band × Language group	**2.74**	**(8, 304)**	**0.006**	**0.067**
Electrode × Word × Frequency band	**9.98**	**(48, 1824)**	**<0.0005**	**0.208**
Word × Frequency band × Language group	0.87	(24, 912)	0.651	0.022
Listening condition × Electrode × Word × Language group	**1.93**	**(12, 456)**	**0.029**	**0.048**
Listening condition × Electrode × Frequency band × Language group	0.78	(8, 304)	0.621	0.020
Listening condition × Electrode × Word × Frequency band	**6.54**	**(48, 1824)**	**<0.0005**	**0.147**
Listening condition × Word × Frequency band × Language group	0.93	(24, 912)	0.562	0.024
Electrode × Word × Frequency band × Language group	0.68	(48, 1824)	0.954	0.018
Listening condition × Electrode × Word × Frequency band × Language group	0.80	(48, 1824)	0.832	0.021

Significant main effects and interactions are bolded and trending differences are indicated by italics.

better performance on the selective attention task. The difference between the correlation for bilinguals and the correlation for monolinguals was significant ($z = -2.275$, $p = 0.011$).

Cortical Phase Consistency

Across all cortical frequency bands, there was a large effect of listening condition [$F(1,38) = 33.26$, $p < 0.0005$, $\eta_p^2 = 0.467$, see **Table 2** for all main effects and interactions from this RMANOVA], with active yielding higher cortical phase consistency than passive (active: 0.098 ± 0.011, passive: 0.089 ± 0.012; **Figures 3**, **4**). Notably, the effect of language group was significant [$F(1,38) = 4.646$, $p = 0.038$, $\eta_p^2 = 0.109$]; monolinguals had higher cortical consistency than bilinguals (monolinguals: 0.097 ± 0.011, bilinguals: 0.090 ± 0.009, **Figures 3**, **4**, **7**). There were also main effects of electrode [$F(2,76) = 55.82$, $p < 0.0005$, $\eta_p^2 = 0.595$], word [$F(6,228) = 42.63$, $p < 0.0005$, $\eta_p^2 = 0.529$], and frequency band [$F(4,152) = 206.12$, $p < 0.0005$, $\eta_p^2 = 0.844$].

With respect to the electrode main effect, Pz (0.0859 ± 0.008) had lower cortical consistency than either Fz [0.098 ± 0.014, $t(39) = 10.126$, $p < 0.0005$, $d = 1.599$] or Cz [0.097 ± 0.012, $t(39) = 7.651$, $p < 0.0005$, $d = 1.214$], while Cz and Fz did not differ [$t(39) = 0.945$, $p = 0.350$, $d = 0.159$].

For the main effect of word, phase consistency was highest at 'ready' (0.113 ± 0.023), followed by the call sign (0.101 ± 0.014), 'Go' (0.059 ± 0.017), the color (0.093 ± 0.015), 'to' (0.086 ± 0.010), the number (0.082 ± 0.012), and the lowest consistency was over 'now' (0.079 ± 0.007). The higher consistency for 'ready' was significant compared to each of the remaining six words in the sentence (all t's 4.205 – 10.785, all p's < 0.0005, all d's 0.664 – 1.706), while the call sign also had significantly higher consistency than 'to,' 'color,' the number, and 'now' (all t's 3.389 – 10.583, all p's ≤ 0.0016, all d's 0.537 – 1.670), 'go' had higher cortical consistency than 'to' [$t(39) = 4.237$, $p < 0.0005$, $d = 0.669$], the number [$t(39) = 6.416$, $p < 0.0005$, $d = 1.013$] and 'now' [$t(39) = 7.065$, $p < 0.0005$, $d = 1.109$], 'to' had lower consistency than the color [$t(39) = 3.640$, $p = 0.001$,

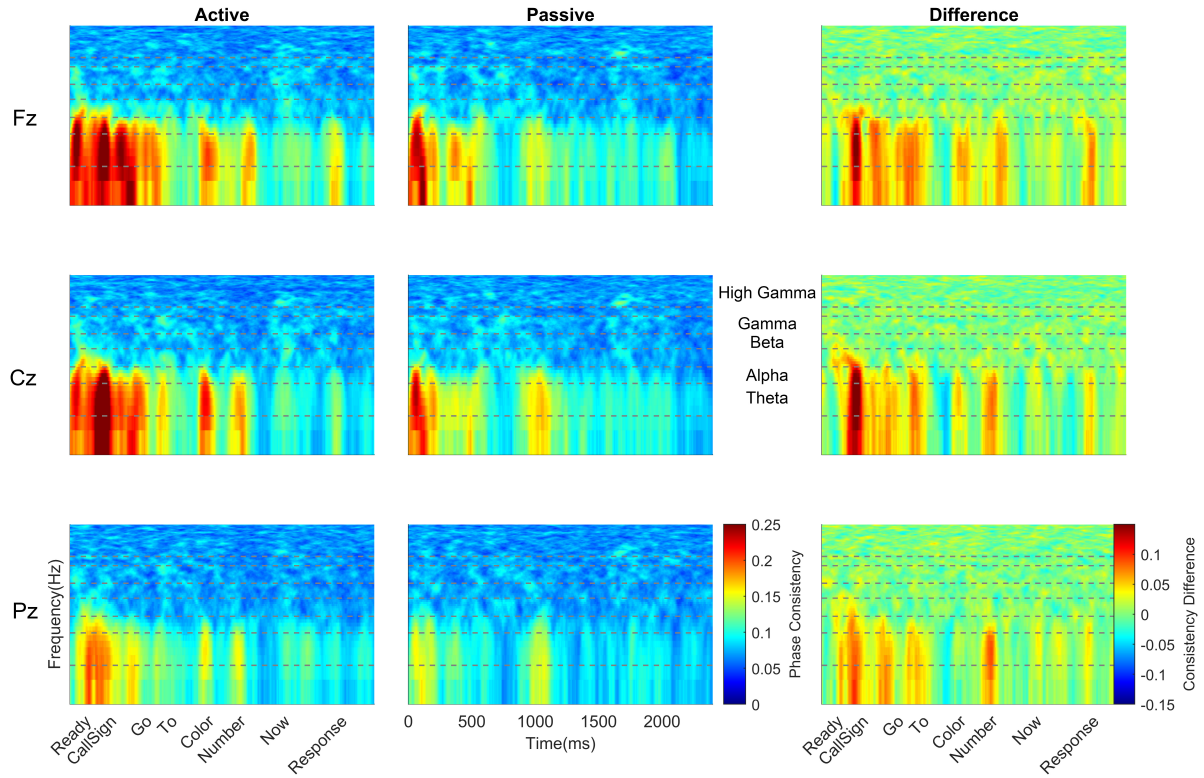

FIGURE 3 | Monolingual cortical phase consistency. Monolingual phase consistency is plotted at Fz **(top)**, Cz **(middle)**, and Pz **(bottom)** for active (left) and passive (center) listening conditions. For these six plots, color represents greater phase consistency, with warmer colors indicating greater consistency and cooler colors representing little to no consistency. The rightmost plots show the difference in phase consistency between active and passive listening conditions, with warmer colors indicating greater consistency during active listening and cooler colors indicating more consistency during passive listening. Monolinguals showed the most cortical consistency during active listening (left column), with the beginning of the sentence showing the highest consistency that declined as the sentence unfolded. This effect was varied across the scalp, being greatest at Fz and lowest at Pz. Active and difference plots align the cortical consistency with the words in the sentence and passive plots provide an indication of time. Dashed lines on all plots identify the borders of the frequency bands, which are labeled between the passive and difference plots for Cz. To visualize the lower frequencies, all *y*-axes are plotted on a log scale.

$d = 0.577$] but higher consistency relative to 'now' [$t(39) = 4.165$, $p < 0.0005$, $d = 0.652$], and the color had higher consistency than the number [$t(39) = 5.364$, $p < 0.0005$, $d = 0.850$] and 'now' [$t(39) = 6.103$, $p < 0.0005$, $d = 0.962$], while the remaining comparisons were not significant (all t's ≤ 2.547, all p's ≥ 0.015).

Considering the main effect of frequency band, as the frequency increased, the phase consistency decreased, such that the greatest consistency was seen over theta (0.141 ± 0.032), followed by alpha (0.118 ± 0.022), beta (0.078 ± 0.009), gamma (0.076 ± 0.007), and high gamma (0.071 ± 0.005). All pairwise frequency-band differences were significant (all t's 5.584 – 14.170, all p's < 0.0005, all d's 0.888 – 2.241) except the difference in consistency between beta and gamma [$t(39) = 1.673$, $p = 0.102$, $d = 0.257$].

To further explore the interaction between listening condition and frequency band, RMANOVAs were run within each frequency band. Within the theta and alpha bands, there was a significant effect of listening condition [theta: $F(1,38) = 42.81$, $p < 0.0005$, $\eta_p^2 = 0.530$; alpha: $F(1,38) = 23.65$, $p < 0.0005$, $\eta_p^2 = 0.384$], while listening condition did not significantly influence phase consistency in the beta, gamma, and high gamma bands (see **Table 3** for all statistics from this analysis). Similar to the overall effect described above, theta and alpha consistency increased during active (theta: 0.145 ± 0.030; alpha: 0.119 ± 0.021), relative to passive (theta: 0.119 ± 0.030; alpha: 0.104 ± 0.021), listening. Also within the theta and alpha bands, there was a significant effect of language group [theta: $F(1,38) = 5.11$, $p = 0.03$, $\eta_p^2 = 0.119$; alpha: $F(1,38) = 5.58$, $p = 0.023$, $\eta_p^2 = 0.128$]. Consistent with the effect described above, monolinguals (theta: 0.141 ± 0.031; alpha: 0.119 ± 0.020) had greater cortical consistency than bilinguals (theta: 0.123 ± 0.019; alpha: 0.105 ± 0.017) over both of these frequency bands (**Figures 3**, **4**). These effects were mirrored in the frequency band by language group interaction, electrode by frequency band by language group interaction, and listening condition, by electrode, by word, by language group interaction, which all demonstrated greater theta and alpha activity for monolinguals relative to bilinguals that was greatest during active listening over Fz and Cz electrodes over the call sign, the words 'go' and 'to' and the number (see **Supplementary Material** results for additional figures and statistics for these analyses and all remaining cortical analyses).

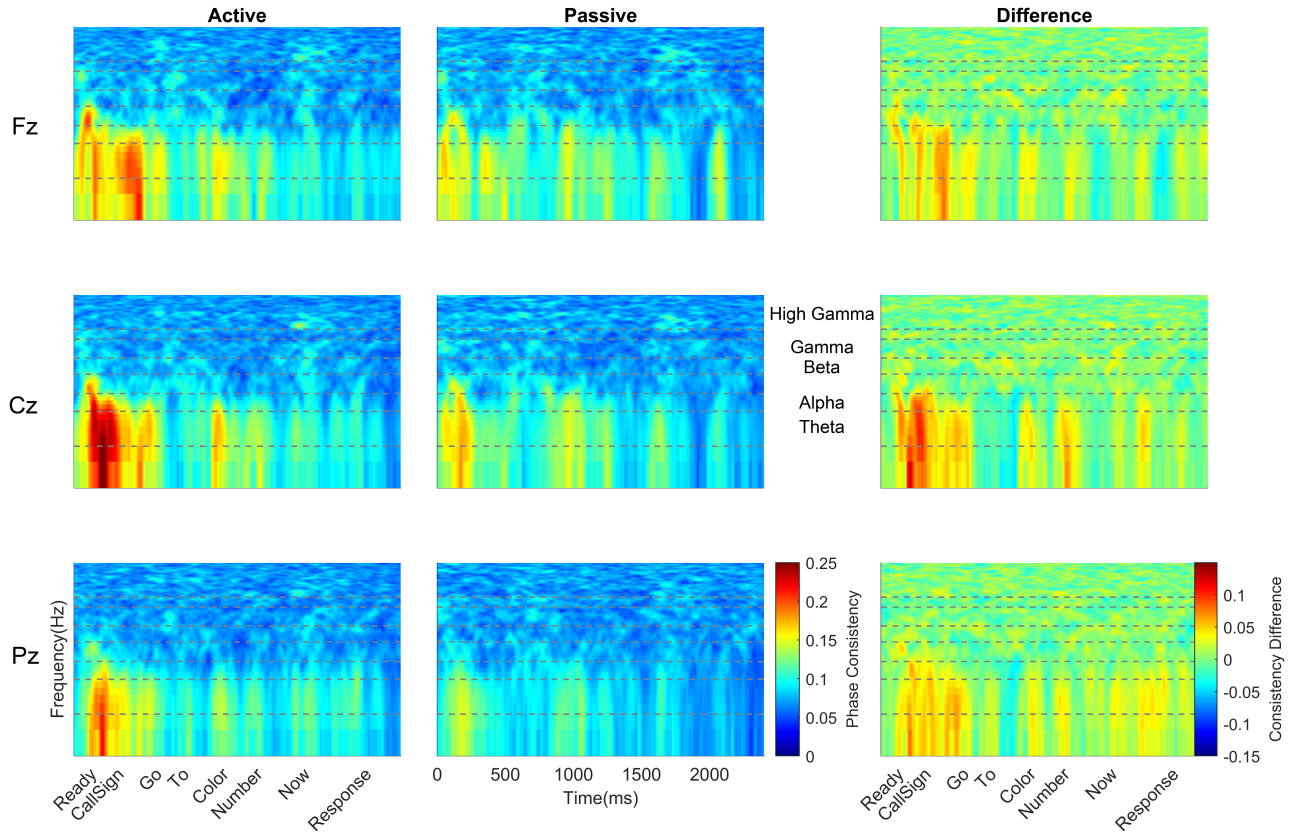

FIGURE 4 | Bilingual cortical phase consistency. Bilingual phase consistency is plotted at Fz **(top)**, Cz **(middle)**, and Pz **(bottom)** for active (left) and passive (center) listening conditions. For these six plots, color represents greater phase consistency, with warmer colors indicating greater consistency and cooler colors representing little to no consistency. The rightmost plots show the difference in phase consistency between active and passive listening conditions, with warmer colors indicating greater consistency during active listening and cooler colors indicating more consistency during passive listening. Compared to monolinguals, bilinguals showed a smaller change in cortical consistency from active to passive listening, driven by a smaller increase in consistency during active listening. However, like monolinguals, bilinguals still had greater cortical consistency during active listening (left column), with consistency highest at the beginning of the sentence and declining as the sentence unfolded. For bilinguals, this effect was greatest at Cz; it did not show the same rostral to caudal distribution seen in monolinguals. Active and difference plots align the cortical consistency with the words in the sentence and passive plots provide an indication of time. Dashed lines on all plots identify the borders of the frequency bands, which are labeled between the passive and difference plots for Cz. To visualize the lower frequencies, all *y*-axes are plotted on a log scale.

Subcortical Phase Consistency

Active versus passive listening also led to differences in subcortical phase consistency [$F(1,38) = 5.60$, $p = 0.023$, $\eta_p^2 = 0.128$; **Table 4**]. However, the effects were in the opposite direction of the changes observed cortically. Whereas *cortical* consistency *increased* during active listening, *subcortical* phase consistency *decreased* during active listening (active: 0.086 ± 0.014; passive: 0.096 ± 0.022; **Figures 5–7**).

Subcortically, there was also a main effect of language group. Across both active and passive listening conditions, bilinguals (0.096 ± 0.016) had greater subcortical phase consistency than monolinguals [0.086 ± 0.008; $F(1,38) = 6.147$, $p = 0.018$, $\eta_p^2 = 0.139$; **Table 4** and **Figures 5–7**].

In addition to the main effects of listening condition and language group, there were also main effects of pitch contour [i.e., male or female talker, $F(1,38) = 80.476$, $p \leq 0.0005$, $\eta_p^2 = 0.679$], electrode [$F(2,76) = 24.703$, $p \leq 0.0005$, $\eta_p^2 = 0.394$], and word [$F(3,114) = 59.183$, $p \leq 0.0005$, $\eta_p^2 = 0.609$]. With respect to the pitch contour, there was higher subcortical consistency to the female pitch (0.100 ± 0.018) relative to the male pitch (0.082 ± 0.011). Interestingly, while Cz and Fz together showed the highest cortical consistency (see above), subcortical consistency was greatest only at Cz. Comparing the three electrodes, Cz (0.094 ± 0.014) had greater subcortical consistency than Fz [0.090 ± 0.013; $t(39) = 6.306$, $p < 0.0005$, $d = 0.985$] and Pz [0.090 ± 0.014; $t(39) = 6.471$, $p < 0.0005$, $d = 1.020$], while Fz and Pz did not differ [$t(39) = 0.734$, $p = 0.467$, $d = 0.103$]. For the words, all word pairs except for 'to' and 'now' [$t(39) = 2.729$, $p = 0.009$, $d = 0.432$] were significantly different. Specifically, 'ready' (0.109 ± 0.025) had greater consistency than 'go' [0.091 ± 0.013, $t(39) = 6.062$, $p < 0.0005$, $d = 0.958$], 'to' [0.085 ± 0.012, $t(39) = 8.526$, $p < 0.0005$, $d = 1.347$], and 'now' [0.081 ± 0.011, $t(39) = 10.575$, $p < 0.0005$, $d = 1.671$]; and 'go' had greater consistency than 'to' [$t(39) = 2.561$, $p = 0.005$, $d = 0.468$] and 'now' [$t(39) = 5.767$, $p < 0.0005$, $d = 0.912$].

TABLE 3 | Cortical analyses within individual frequency bands.

		Theta			Alpha			Beta			Gamma			High Gamma		
	df	F	p	η_p^2	F	p	η_p^2	F	p	η_p^2	F	p	η_p^2	F	p	η_p^2
Listening condition	1, 38	**42.81**	**<0.0005**	**0.530**	**23.65**	**<0.0005**	**0.384**	2.76	0.105	0.068	0.001	0.976	0	0.92	0.345	0.024
Language group	1, 38	**5.11**	**0.030**	**0.119**	**5.58**	**0.023**	**0.128**	0.12	0.732	0.003	*3.17*	*0.083*	*0.077*	1.60	0.214	0.040
Electrode	2, 76	**35.58**	**<0.0005**	**0.484**	**56.79**	**<0.0005**	**0.599**	**11.07**	**<0.0005**	**0.226**	**22.63**	**<0.0005**	**0.373**	**5.34**	**<0.0005**	**0.123**
Word	6, 228	**29.08**	**<0.0005**	**0.434**	**35.70**	**<0.0005**	**0.484**	**10.53**	**<0.0005**	**0.217**	**17.19**	**<0.0005**	**0.311**	**14.09**	**<0.0005**	**0.270**
Listening condition × Language group	1, 38	1.16	0.288	0.030	*3.05*	*0.089*	*0.074*	0.66	0.420	0.017	0.19	0.669	0.005	0.03	0.863	0.001
Listening condition × Electrode	2, 76	1.21	0.305	0.031	**5.16**	**0.008**	**0.120**	1.06	0.351	0.027	0.03	0.966	0.001	2.36	0.102	0.058
Listening condition × Word	6, 228	**5.91**	**<0.0005**	**0.135**	**11.75**	**<0.0005**	**0.219**	**4.28**	**<0.0005**	**0.101**	0.26	0.957	0.007	0.39	0.888	0.010
Electrode × Language Group	2, 76	*2.93*	*0.059*	*0.072*	1.59	0.211	0.040	0.91	0.406	0.023	2.05	0.135	0.051	0.09	0.917	0.002
Electrode × Word	12, 456	**3.11**	**<0.0005**	**0.076**	**6.20**	**<0.0005**	**0.140**	**2.20**	**0.011**	**0.055**	*1.64*	*0.077*	*0.041*	*1.68*	*0.069*	*0.042*
Word × Language group	6, 228	0.99	0.436	0.025	1.06	0.387	0.027	0.20	0.977	0.005	0.46	0.840	0.012	*2.09*	*0.056*	*0.052*
Listening condition × Electrode × Language group	2, 76	*2.39*	*0.099*	*0.059*	1.90	0.157	0.048	0.12	0.885	0.003	0.12	0.889	0.003	0.14	0.866	0.004
Listening condition × Word × Language group	6, 228	0.91	0.492	0.023	0.61	0.721	0.016	0.77	0.592	0.020	0.28	0.971	0.006	0.58	0.743	0.015
Listening condition × Electrode × Word	12, 456	**1.84**	**0.040**	**0.046**	1.39	0.168	0.035	0.82	0.632	0.021	1.38	0.172	0.035	0.68	0.776	0.017
Electrode × Word × Language group	12, 456	1.09	0.363	0.028	0.66	0.793	0.017	0.65	0.795	0.017	*1.63*	*0.081*	*0.041*	1.28	0.227	0.033
Listening condition × Electrode × Word × Language group	12, 456	0.75	0.632	0.019	0.78	0.671	0.020	0.87	0.580	0.022	**1.93**	**0.029**	**0.048**	*1.60*	*0.088*	*0.040*

Significant main effects and interactions are bolded and trending differences are indicated by italics.

In addition to these main effects, there were a number of interactions, whose results are in line with those detailed above and are described fully in the **Supplementary Material**. Briefly, we observed that the greatest differences between active and passive listening were in response to the female pitch contour and that these effects were largest earlier in the sentence, such that the greatest consistency was in response to the female 'ready' during passive listening.

Cortical – Subcortical Comparisons

Effects of Language Experience

To understand how cortical and subcortical processing work in tandem during active and passive listening, and whether language experience influences the interaction between cortical and subcortical processing, we compared cortical and subcortical phase consistency across bilinguals and monolinguals. There were no main effects of listening condition [$F(1,38) = 0.224$, $p = 0.639$, $\eta_p^2 = 0.006$], auditory center {i.e., cortical vs. subcortical, [$F(1,38) = 0.016$, $p = 0.901$, $\eta_p^2 = 0$]}, or language group [$F(1,38) = 0.621$, $p = 0.435$, $\eta_p^2 = 0.016$]. Nor were there a listening condition by language group interaction ($F(1,38) = 0.705$, $p = 0.406$, $\eta_p^2 = 0.018$) or listening condition by auditory level by language group three-way interaction [$F(1,38) = 0.004$, $p = 0.952$, $\eta_p^2 = 0$). However, both the auditory center by language group [$F(1,38) = 11.785$, $p = 0.001$, $\eta_p^2 = 0.237$] and listening condition by auditory center [$F(1,38) = 17.999$, $p < 0.0005$, $\eta_p^2 = 0.321$] interactions were significant. These differences were driven by (1) greater cortical consistency for monolinguals, (2) greater subcortical consistency for bilinguals, (3) greater cortical consistency during active listening, and (4) greater subcortical consistency during passive listening. Interestingly, these effects resulted in matched levels of cortical and subcortical auditory consistency for bilinguals during *active* listening, caused by a reduction in subcortical consistency and an increase in cortical consistency (**Figure 7**, red lines). In contrast, monolinguals' cortical and subcortical consistency were matched during *passive* listening, driven by a reduction in

TABLE 4 | Subcortical main effects and interactions for a 2 (language group: monolingual, bilingual) × 2 (listening condition: active, passive) × 3 (electrode: Fz, Cz, Pz) × 2 (pitch contour: male talker, female talker) × 4 (word: 'ready,' 'go,' 'to,' 'now') RMANOVA.

	F	df	*p*	η_p^2
Listening condition	**5.60**	**(1, 38)**	**0.023**	**0.128**
Language group	**6.15**	**(1, 38)**	**0.018**	**0.139**
Pitch contour	**80.48**	**(1, 38)**	**<0.0005**	**0.679**
Electrode	**24.70**	**(2, 76)**	**<0.0005**	**0.394**
Word	**59.18**	**(3, 114)**	**<0.0005**	**0.609**
Listening condition × Language group	0.18	(1, 38)	0.671	0.005
Pitch contour × Language group	0.29	(1, 38)	0.592	0.008
Electrode × Language group	0.47	(2, 76)	0.629	0.012
Word × Language group	**3.28**	**(3, 114)**	**0.024**	**0.079**
Listening condition × Pitch contour	**10.30**	**(1, 38)**	**0.003**	**0.213**
Listening condition × Electrode	0.01	(2, 76)	0.987	0
Listening condition × Word	**5.01**	**(3, 114)**	**0.003**	**0.117**
Pitch contour × Electrode	**17.29**	**(2, 76)**	**<0.0005**	**0.313**
Pitch contour × Word	**38.25**	**(3, 114)**	**<0.0005**	**0.502**
Word × Electrode	**4.11**	**(6, 228)**	**0.001**	**0.098**
Listening condition × Pitch contour × Language group	1.01	(1, 38)	0.321	0.026
Listening condition × Word × Language group	0.82	(3, 114)	0.486	0.021
Listening condition × Pitch contour × Word	**5.13**	**(3, 114)**	**0.002**	**0.119**
Listening condition × Electrode × Language group	2.24	(2, 76)	0.114	0.056
Listening condition × Pitch contour × Electrode	0.44	(2, 76)	0.645	0.011
Listening condition × Word × Electrode	0.17	(6, 228)	0.984	0.005
Pitch contour × Word × Language group	1.98	(3, 114)	0.121	0.050
Pitch contour × Electrode × Language group	1.90	(2, 76)	0.156	0.048
Pitch contour × Word × Electrode	**3.09**	**(6, 228)**	**0.006**	**0.075**
Word × Electrode × Language group	0.77	(6, 228)	0.595	0.020
Listening condition × Pitch contour × Word × Language group	0.29	(3, 114)	0.835	0.007
Listening condition × Pitch contour × Electrode × Language group	0.70	(2, 76)	0.498	0.018
Listening condition × Word × Electrode × Language group	0.91	(6, 228)	0.487	0.023
Listening condition × Pitch contour × Word × Electrode	**2.83**	**(6, 228)**	**0.011**	**0.069**
Pitch contour × Word × Electrode × Language group	1.09	(6, 228)	0.372	0.028
Listening condition × Pitch contour × Word × Electrode × Language group	0.67	(6, 228)	0.674	0.017

Significant main effects and interactions are bolded.

cortical consistency and an increase in subcortical consistency (**Figure 7**, black lines).

Comparing High Gamma and the Male Pitch Contour

To determine whether neural consistency over high gamma and in response to the male pitch contour reflected different sources, we analyzed these responses to determine if they were statistically different, if they were influenced differently by listening conditions, and if they showed different patterns across electrodes and/or words. We found that there was a difference between the responses and that they patterned differently across electrodes and words over the two listening conditions {main effect of response: $F(1,39) = 46.48$, $p < 0.0005$, $\eta_p^2 = 0.544$, response × listening condition × electrode × word interaction [$F(6,234) = 2.22$, $p = 0.042$, $\eta_p^2 = 0.054$, **Figure 8** and see **Table 5** for additional statistics]}. Specifically, high gamma consistency did not differ between the two listening conditions and, during both active and passive listening, was lower in consistency than the male pitch contour across the four words. In contrast, the male pitch contour showed an attention effect consistent with the effect seen for the female pitch contour: consistency to both contours increased during passive listening.

DISCUSSION

Bilinguals previously were shown to have enhanced inhibitory control (Bialystok, 2011, 2015) and subcortical encoding of the F0 of speech (Krizman et al., 2012; Skoe et al., 2017), processes that are fundamental to understanding speech in noise. Despite these advantages, bilinguals perform more poorly on clinical tests of speech-in-noise recognition (Shi, 2010, 2012; Lucks Mendel and Widner, 2016; Krizman et al., 2017; Skoe and Karayanidi, 2019). In this study, by assessing bilinguals and monolinguals on a selective attention task, a type of speech-in-noise task that calls upon auditory processing and executive control, we could determine whether bilingual cognitive and sensory enhancements have benefits for everyday listening situations. We find that although monolinguals and bilinguals performed similarly on the behavioral component of the selective

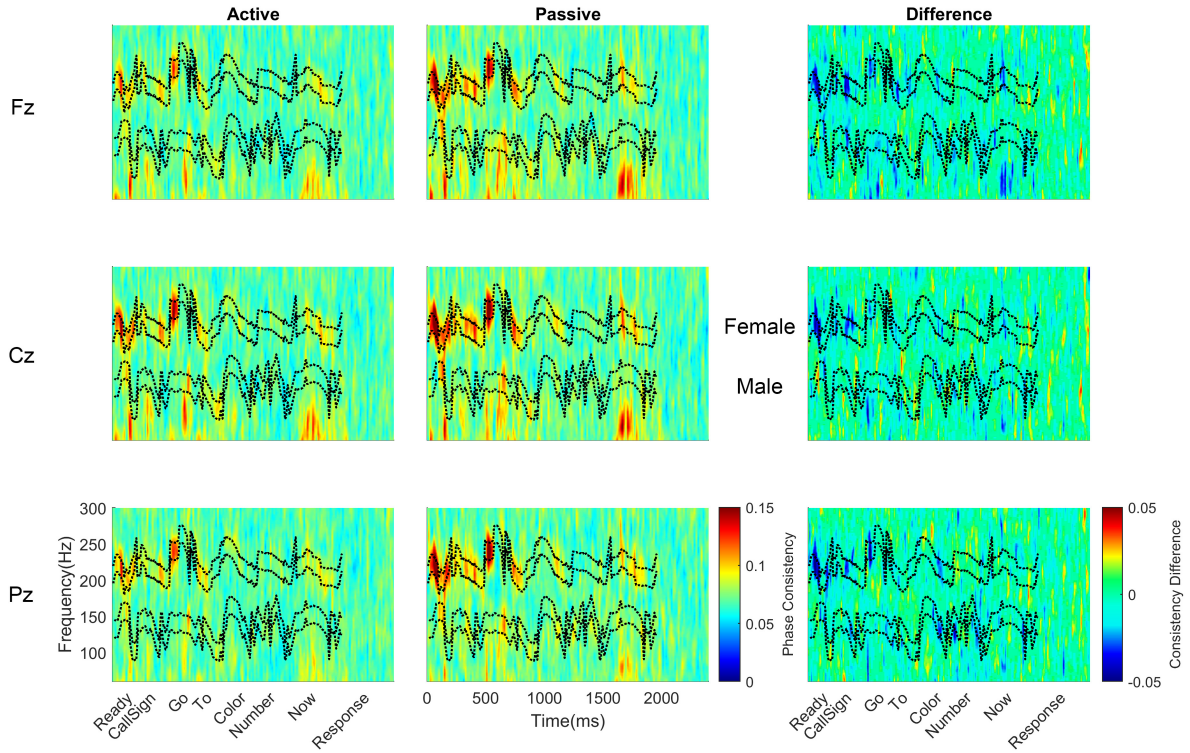

FIGURE 5 | Monolingual subcortical phase consistency. Monolingual phase consistency is plotted at Fz **(top)**, Cz **(middle)**, and Pz **(bottom)** for active (left) and passive (center) listening conditions. For these six plots, color represents greater phase consistency, with warmer colors indicating greater consistency and cooler colors representing little to no consistency. The rightmost plots show the difference in phase consistency between active and passive listening conditions, with warmer colors indicating greater consistency during active listening and cooler colors indicating more consistency during passive listening. The top two black dashed lines in each plot indicate the female pitch contour ± 10 Hz and the bottom two dashed lines indicate the male pitch contour ± 10 Hz (labeled between the passive and difference plots for Cz). Note that the regions that have multiple words (e.g., number) have no phase consistency while phase consistency is evident over the words that are identical across trials (i.e., 'ready,' 'go,' 'to,' and 'now'). Monolinguals' subcortical phase consistency decreased during active listening, in contrast to the effects of active listening on cortical phase consistency.

attention task, the groups differed in the neural and cognitive processes engaged to perform this task. Specifically, bilinguals demonstrated enhanced subcortical phase-locking to the pitch contours of the talkers, while monolinguals had enhanced cortical phase consistency during active listening, particularly over the theta and alpha frequency bands. Additionally, a relationship between performance on the selective attention task and the inhibitory control test was seen only in bilinguals. Together, these results suggest that bilinguals utilize inhibitory control and enhanced subcortical auditory processing in real-world listening situations and are consistent with the hypothesis that bilingualism leads to mechanistic differences in how the brain engages with sound (Abutalebi et al., 2011; Ressel et al., 2012; Costa and Sebastián-Gallés, 2014; García-Pentón et al., 2014; Krizman et al., 2017).

Monolinguals' and bilinguals' equivalent performance on the selective attention task may seem inconsistent with previous literature showing that bilinguals perform more poorly than monolinguals on tests of speech-in-noise recognition (Mayo et al., 1997; Shi, 2010, 2012). However in the present study, the consistent structure of the sentences across trials, together with the limited number of words that could potentially appear in the sentence, likely limited the linguistic complexity, and thus, the linguistic processing demands of the task. Previous findings of a bilingual disadvantage, especially for early-acquiring, highly proficient bilinguals similar to the ones tested in the current study, used sentences that are semantically and syntactically correct, but unrestricted in their content or word choice (Mayo et al., 1997; Shi, 2010, 2012). This open-endedness increases linguistic processing demands. When the target is restricted, such as when a target word is embedded in a carrier phrase, early, high-proficiency bilinguals have been reported to perform equivalently to their monolingual peers (Krizman et al., 2017). Additionally, differences in speech-in-noise recognition between monolinguals and bilinguals are starkest when the sentences contain semantic context that can be used to 'fill in the gaps' (Bradlow and Bent, 2002; Bradlow and Alexander, 2007). While monolinguals are able to benefit from the semantic context contained within a degraded sentence, bilinguals benefit less from this context (Mayo et al., 1997; Bradlow and Alexander, 2007). Given that the sentences used here did not contain any semantic context to aid in disambiguation between the target and irrelevant talker, top–down semantic knowledge could not aid monolinguals' performance on this task. Thus, the linguistic

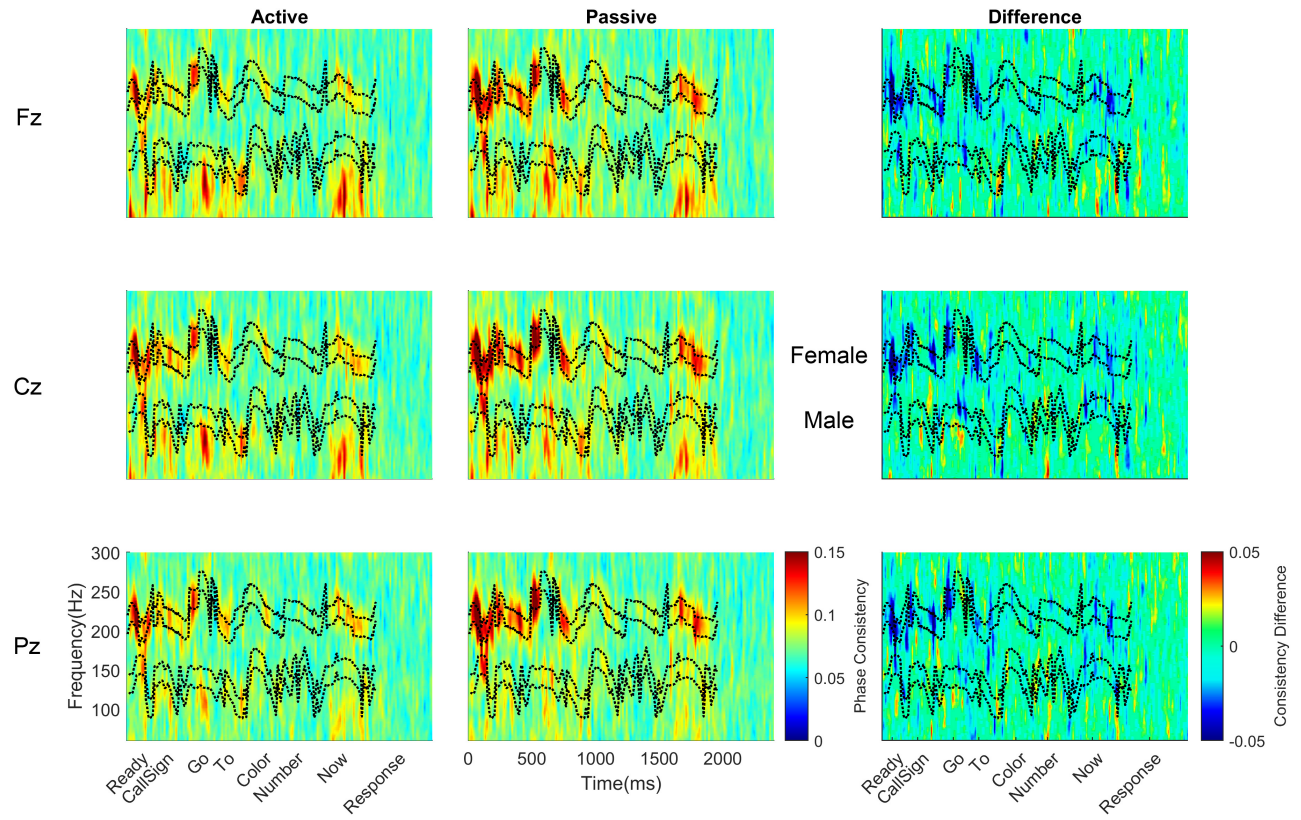

FIGURE 6 | Bilingual subcortical phase consistency. Bilingual phase consistency is plotted at Fz **(top)**, Cz **(middle)**, and Pz **(bottom)** for active (left) and passive (center) listening conditions. For these six plots, color represents greater phase consistency, with warmer colors indicating greater consistency and cooler colors representing little to no consistency. The rightmost plots show the difference in phase consistency between active and passive listening conditions, with warmer colors indicating greater consistency during active listening and cooler colors indicating more consistency during passive listening. The top two black dashed lines in each plot indicate the female pitch contour $\pm$10 Hz and the bottom two dashed lines indicate the male pitch contour $\pm$10 Hz (labeled between the passive and difference plots for Cz). Note that the regions that have multiple words (e.g., number) have no phase-locking while phase-locking consistency is evident over the words that are identical across trials (i.e., 'ready,' 'go,' 'to,' and 'now'). Similar to monolinguals, active listening led to a decline in phase consistency in the subcortical response. However, across both passive and active listening conditions, bilinguals had higher subcortical phase consistency than monolinguals.

cues that lead to performance differences between bilinguals and monolinguals on speech-in-noise tasks were unavailable here.

Although the groups performed similarly, the neural analyses suggest that bilinguals and monolinguals utilize different listening strategies. Both groups showed an increase in cortical phase consistency during active listening, coupled with a decrease in subcortical phase consistency. However, relative to one another, monolinguals had greater cortical consistency, especially during active listening, while bilinguals had greater subcortical consistency, especially during passive listening.

When looking at the differences in cortical consistency between the two groups, the effect was concentrated over the theta and alpha bands. Given that the energy of the sentence envelopes used here was concentrated at 4 Hz, which is within the 3–7 Hz theta band, the theta differences are likely to reflect more consistent cortical tracking of the stimulus envelope during active listening and in monolinguals. In addition to more consistent tracking of the stimulus envelope, it is possible that the theta, and potentially alpha, differences are driven by greater cortical evoked potentials to each word in the sentence, similar to what has been demonstrated previously in bilinguals and monolinguals (Astheimer et al., 2016). Because the stimuli were designed to facilitate the subcortical FFR recording, the envelope and the cortical potentials overlap in time and frequency, and so it is difficult to disentangle the contribution of each. Another potential source of the enhanced alpha activity is the greater load that is placed on cognitive processing, particularly working memory, during active listening, consistent with previous findings that increases in alpha synchrony during a task are related to the working memory requirements of that task (Jensen et al., 2002; Jensen and Hanslmyar, 2020). If the cortical differences are tied to executive functions, the reduced cortical consistency in bilinguals may result from decreased recruitment of cortical brain regions involved in this task, similar to the reduction in activation of lateral frontal cortex and anterior cingulate cortex when doing a complex task that requires conflict monitoring, in bilinguals relative to monolinguals (Abutalebi et al., 2011; Gold et al., 2013; Costa and Sebastián-Gallés, 2014). That is, bilinguals may require less cortical processing

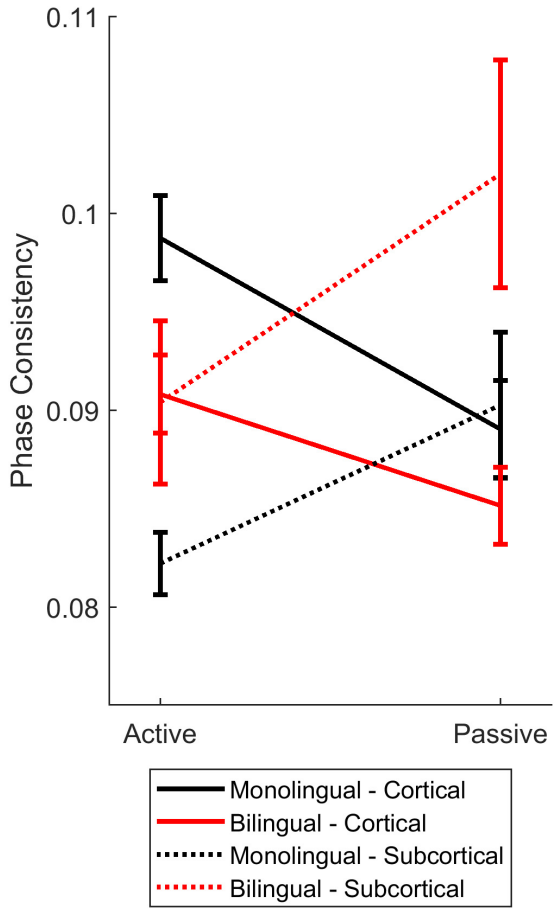

FIGURE 7 | Cortical-subcortical comparisons in bilinguals and monolinguals. Differences in levels of cortical (solid lines) and subcortical (dashed lines) consistency for monolinguals (black) and bilinguals (red) during active (left) and passive (right) listening.

to perform these tasks at a level that is similar to, or better than, monolinguals.

The enhanced subcortical pitch encoding for bilinguals is consistent with their previously reported F0 enhancement (Krizman et al., 2012; Skoe et al., 2017) and suggests that bilinguals are acutely tuned-in to the acoustic features, namely the F0, of a talker and use that cue to understand speech in challenging listening situations. In a competing-talkers environment, the F0 can be used to track the target talker, and was likely a useful cue here given the difference in F0 between the two talkers. The finding that active listening decreases subcortical consistency is supported by earlier work in animals finding that active listening leads to diminished responses in the inferior colliculus, the predominant contributor of the FFR (Chandrasekaran and Kraus, 2010; Bidelman, 2015, 2018; Coffey et al., 2019; White-Schwoch et al., 2019). However, a very recent study in humans found that active listening leads to increases in the FFR (Price and Bidelman, 2021). The different outcomes of these studies raise an intriguing possibility: that corticofugal tuning of the subcortical response can lead to differences in how attention effects manifest in the scalp-recorded FFR. In addition to the numerous ascending projections from the ear to the brain, there exists an even larger population of descending projections connecting the various auditory centers between the brain and the ear (Malmierca and Ryugo, 2011; Malmierca, 2015). These pathways regulate the incoming signal to meet the demands of the task, which can result in inhibition or amplification of the incoming signal depending on its task relevance (Malmierca et al., 2009, 2019; Parras et al., 2017; Ito and Malmierca, 2018). Future studies should investigate the task dependency on subcortical attention effects.

Bilinguals appear to also call upon their inhibitory control abilities during active listening, as performance on the inhibitory control and selective attention tasks was related only in this language group. We hypothesize that auditory processing and inhibitory control work in tandem to compensate, at least partly, for the greater demands that bilingualism places on language processing. This hypothesis is supported by previous findings of relationships between inhibitory control and auditory processing that are specific to bilinguals (Blumenfeld and Marian, 2011; Krizman et al., 2012, 2014; Marian et al., 2018). In contrast to bilinguals, monolinguals have greater experience with the target language and can rely more heavily on linguistic cues (e.g., linguistic context) to understand speech, particularly in difficult listening conditions (Cooke et al., 2008; Lecumberri et al., 2011; Mattys et al., 2012; Krizman et al., 2017; Strori et al., 2020). Because bilinguals are less able to benefit from these cues, we propose that they rely on non-linguistic processes, specifically sensory encoding and executive control, to overcome this disadvantage.

Although they used inhibitory control differently, bilinguals and monolinguals performed equivalently on the inhibitory control task. The similar performance between language groups contrasts with previous studies showing an inhibitory control advantage for bilinguals (Bialystok and Martin, 2004; Bialystok et al., 2005; Carlson and Meltzoff, 2008; Krizman et al., 2012, 2014); although, the evidence for a bilingual inhibitory control advantage has been equivocal, with other studies reporting that no such advantage exists (Bialystok et al., 2015; de Bruin et al., 2015; Paap et al., 2015; Dick et al., 2019). Interestingly, many (but not all) studies that do find an advantage tend to find it when comparing bilingual and monolingual participants at the ends of the lifespan (i.e., young children and older adults) while many that do not find an advantage have looked for performance differences in young adults. This may suggest that bilinguals and monolinguals eventually reach the same level of inhibitory control abilities but that bilinguals mature to this level at a faster rate (and decline from this level more slowly later in life). Given that inhibitory control is malleable with other enriching life experiences, such as music training (Moreno et al., 2014; Slater et al., 2018) or sports participation (Lind et al., 2019; Hagyard et al., 2021), it may be more difficult to isolate the bilingual enhancement when looking across individuals from different backgrounds, especially with increasing age and enrichment. If bilinguals and monolinguals use this skill differently in everyday settings, it could explain the re-emergence

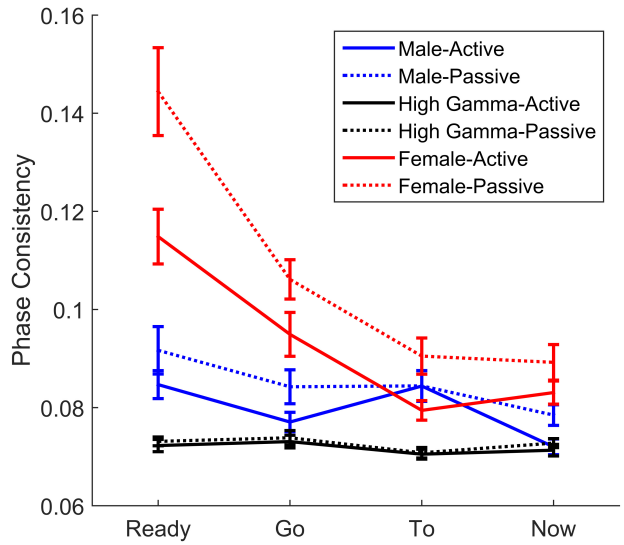

FIGURE 8 | Comparisons of high gamma, male and female pitch contours across words and listening condition. While cortical high gamma phase consistency did not differ between conditions, both the subcortical phase consistency in response to the male and female pitch contours decreased during active listening.

TABLE 5 | RMANOVA comparisons of high gamma and male pitch contour consistency over active and passive listening conditions.

	F	df	*p*	η_p^2
Response	**46.48**	**(1, 39)**	**<0.0005**	**0.544**
Listening condition	**2.23**	**(1, 39)**	**0.143**	**0.054**
Electrode	**10.82**	**(2, 78)**	**<0.0005**	**0.217**
Word	**9.24**	**(3, 117)**	**<0.0005**	**0.192**
Response × Electrode	**3.39**	**(2, 78)**	**0.039**	**0.080**
Response × Word	**9.77**	**(3, 117)**	**<0.0005**	**0.200**
Listening condition × Response	2.35	(1, 39)	0.133	0.057
Listening condition × Electrode	0.50	(2, 78)	0.611	0.013
Listening condition × Word	1.52	(3, 117)	0.214	0.037
Electrode × Word	0.98	(6, 234)	0.437	0.025
Listening condition × Response × Word	1.70	(3, 117)	0.171	0.042
Electrode × Listening condition × Response	1.98	(2, 78)	0.145	0.048
Electrode × Listening condition × Word	0.47	(6, 234)	0.830	0.012
Electrode × Response × Word	**2.73**	**(6, 234)**	**0.014**	**0.065**
Response × Listening condition × Electrode × Word	**2.22**	**(6, 234)**	**0.042**	**0.054**

Significant main effects and interactions are bolded.

of a bilingual inhibitory-control advantage in older adults (Bialystok et al., 2005).

Separating High Gamma and Pitch-Contour Tracking

There is general consensus that the response to the female voice is subcortical in origin, given the higher frequency of that voice (∼220 Hz) and that cortical phase-locking limitations preclude reliable firing to this frequency (Aiken and Picton, 2008; Akhoun et al., 2008; Bidelman, 2018). However, there is still some debate about the origins of the response to the male voice in this study, given the overlap between the high gamma frequency range and the pitch of the male talker. Nevertheless, these responses are presumed to originate from distinct sources (Edwards et al., 2005; Chandrasekaran and Kraus, 2010; Mesgarani and Chang, 2012; White-Schwoch et al., 2019). To determine whether these responses indeed reflect distinct sources, we compared them to see if they differed from one another and if they were affected differently by different stimulus and protocol parameters. We did observe differences between the high gamma and pitch contour responses. We found that high gamma consistency did not differ between the two listening conditions and that across the two conditions and the four words ('ready,' 'go,' 'to,' 'now'), it was lower than the consistency to the male pitch contour. Similar to the response to the female pitch contour, consistency of the response to the male pitch contour increased during passive listening. Given that the female pitch contour is above the cortical phase-locking limits but within subcortical phase-locking limits, it is presumed to arise from subcortical sources (Liang-Fa et al., 2006; Chandrasekaran and Kraus, 2010; Coffey et al., 2016, 2019). The difference between the high gamma and male pitch contour consistency effects, together with the similar effects of listening condition on the male and female pitch contours suggest that these pitch contour responses arise from similar subcortical sources (White-Schwoch et al., 2019). Furthermore, a 10 ms-lag was used to analyze pitch-tracking consistency to align the analyses with the temporal lag between the sound reaching the ear and reaching the brainstem, a much faster lag than that seen for the cortex, which arises ∼40 ms after the stimulus is first heard (Langner and Schreiner, 1988; Liegeois-Chauvel et al., 1994; Coffey et al., 2016). Although the clearest way to classify the activity as distinctly cortical and subcortical would be through either source localization or direct simultaneous recordings from the regions of interest in an animal model, these methods used to analyze these responses and the differences found between high-gamma and pitch-tracking activity support the hypothesis that these responses arise from distinct cortical and subcortical sources, respectively. We acknowledge, however, that there still is some debate about the origins of the FFR, when evoked at frequencies around those of the male speaker, as some MEG studies suggest that there is a cortical contribution to the FFR at that frequency range in addition to a larger subcortical contribution, which may influence findings of attention effects on FFR (Bidelman, 2018; Coffey et al., 2019; Hartmann and Weisz, 2019).

CONCLUSION

In conclusion, results from this study are consistent with the hypothesis that bilinguals utilize their cognitive and sensory enhancements for active listening. We found that monolingual and bilingual adolescents and young adults differed

in the neural and cognitive processes engaged to perform a selective attention task, yet performed similarly on the task. Specifically, although both groups showed an increase in cortical phase consistency during active listening, coupled with a decrease in subcortical phase consistency, relative to one another, monolinguals had greater cortical consistency, especially during active listening, while bilinguals had greater subcortical consistency. Also, bilinguals showed a relationship between performance on the inhibitory control and selective attention tests, while monolinguals did not. The neural findings highlight an interesting distinction between online and lifelong modulation of midbrain auditory processing. The bilingual enhancement coupled with the active-listening suppression suggest that different mechanisms underlie short-term and long-term changes in subcortical auditory processing.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Northwestern University Institutional Review Board. Written informed consent to participate in this study was provided by participants 18 and older while informed written assent was given by adolescents younger than 18 and consent provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

JK, AT, TN, and NK: conceptualization, methodology, writing – review and editing. JK and AT: investigation. JK: project administration and writing – original draft. JK, AT, and TN: visualization. JK and NK: funding acquisition. NK: supervision. All authors contributed to the article and approved the submitted version.

FUNDING

This research was funded by F31 DC014221 to JK, the Mathers Foundation, and the Knowles Hearing Center, Northwestern University to NK.

ACKNOWLEDGMENTS

The authors thank all of the participants and their families. The authors are also grateful to Travis White-Schwoch and Erika Skoe, for helpful discussions about the study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2021.717572/full#supplementary-material

REFERENCES

Abutalebi, J., Della Rosa, P. A., Green, D. W., Hernandez, M., Scifo, P., Keim, R., et al. (2011). Bilingualism tunes the anterior cingulate cortex for conflict monitoring. *Cereb. Cortex* 22, 2076–2086. doi: 10.1093/cercor/bhr287

Abutalebi, J., and Green, D. W. (2007). Bilingual language production: the neurocognition of language representation and control. *J. Neurolinguistics* 20, 242–275. doi: 10.1016/j.jneuroling.2006.10.003

Aiken, S. J., and Picton, T. W. (2008). Envelope and spectral frequency-following responses to vowel sounds. *Hear. Res.* 245, 35–47. doi: 10.1016/j.heares.2008.08.004

Akhoun, I., Gallego, S., Moulin, A., Menard, M., Veuillet, E., Berger-Vachon, C., et al. (2008). The temporal relationship between speech auditory brainstem responses and the acoustic pattern of the phoneme /ba/ in normal-hearing adults. *Clin. Neurophysiol.* 119, 922–933.

Alain, C., and Woods, D. L. (1999). Age-related changes in processing auditory stimuli during visual attention: evidence for deficits in inhibitory control and sensory memory. *Psychol. Aging* 14, 507–19.

Anderson, S., Skoe, E., Chandrasekaran, B., Zecker, S., and Kraus, N. (2010). Brainstem correlates of speech-in-noise perception in children. *Hear. Res.* 270, 151–157. doi: 10.1016/j.heares.2010.08.001

Astheimer, L. B., Berkes, M., and Bialystok, E. (2016). Differential allocation of attention during speech perception in monolingual and bilingual listeners. *Lang. Cogn. Neurosci.* 31, 196–205.

Baddeley, A. (2003). Working memory and language: an overview. *J. Commun. Disord.* 36, 189–208.

Bialystok, E. (2009). Bilingualism: the good, the bad, and the indifferent. *Biling. Lang. Cogn.* 12, 3–11.

Bialystok, E. (2011). Reshaping the mind: the benefits of bilingualism. *Can. J. Exp. Psychol.* 65, 229–235. doi: 10.1037/a0025406

Bialystok, E. (2015). Bilingualism and the development of executive function: the role of attention. *Child Dev. Pers.* 9, 117–121. doi: 10.1111/cdep.12116

Bialystok, E., Kroll, J. F., Green, D. W., MacWhinney, B., and Craik, F. I. (2015). Publication bias and the validity of evidence what's the connection? *Psychol. Sci.* 26, 944–6.. doi: 10.1177/0956797615573759

Bialystok, E., and Martin, M. M. (2004). Attention and inhibition in bilingual children: evidence from the dimensional change card sort task. *Dev. Sci.* 7, 325–339. doi: 10.1111/j.1467-7687.2004.00351.x

Bialystok, E., Martin, M. M., and Viswanathan, M. (2005). Bilingualism across the lifespan: the rise and fall of inhibitory control. *Int. J. Biling.* 9, 103–119.

Bialystok, E., and Viswanathan, M. (2009). Components of executive control with advantages for bilingual children in two cultures. *Cognition* 112, 494–500.

Bidelman, G. M. (2015). Multichannel recordings of the human brainstem frequency-following response: scalp topography, source generators, and distinctions from the transient ABR. *Hear. Res.* 323, 68–80. doi: 10.1016/j.heares.2015.01.011

Bidelman, G. M. (2018). Subcortical sources dominate the neuroelectric auditory frequency-following response to speech. *Neuroimage* 175, 56–69. doi: 10.1016/j.neuroimage.2018.03.060

Bird, J., and Darwin, C. (1998). "Effects of a difference in fundamental frequency in separating two sentences," in *Psychophysical and Physiological Advances in Hearing*, eds A. R. Palmer, A. Rees, A. Q. Summerfield, and R. Meddis, (London: Whurr), 263–269.

Blumenfeld, H. K., and Marian, V. (2011). Bilingualism influences inhibitory control in auditory comprehension. *Cognition* 118, 245–257. doi: 10.1016/j.cognition.2010.10.012

Bolia, R. S., Nelson, W. T., Ericson, M. A., and Simpson, B. D. (2000). A speech corpus for multitalker communications research. *J. Acoust. Soc. Am.* 107, 1065–6. doi: 10.1121/1.428288

Bradlow, A. R., and Alexander, J. A. (2007). Semantic and phonetic enhancements for speech-in-noise recognition by native and non-native listeners. *J. Acoust. Soc. Am.* 121, 2339–49. doi: 10.1121/1.2642103

Bradlow, A. R., and Bent, T. (2002). The clear speech effect for non-native listeners. *J. Acoust. Soc. Am.* 112, 272–84. doi: 10.1121/1.1487837

Bregman, A. S., Liao, C., and Levitan, R. (1990). Auditory grouping based on fundamental frequency and formant peak frequency. *Can. J. Psychol.* 44, 400–13. doi: 10.1037/h0084255

Carlson, S. M., and Meltzoff, A. N. (2008). Bilingual experience and executive functioning in young children. *Dev. Sci.* 11, 282–298.

Casseday, J. H., Fremouw, T., and Covey, E. (2002). "The inferior colliculus: a hub for the central auditory system," in *Integrative Functions in the Mammalian Auditory Pathway*, eds D. Oertel, R. R. Fay, and A. N. Popper, (New York: Springer), 238–318. doi: 10.1007/978-1-4757-3654-0_7

Cervenka, M. C., Nagle, S., and Boatman-Reich, D. (2011). Cortical high-gamma responses in auditory processing. *Am. J. Audiol.* 20, 171–180. doi: 10.1044/1059-0889(2011/10-0036)

Chandrasekaran, B., and Kraus, N. (2010). The scalp-recorded brainstem response to speech: neural origins and plasticity. *Psychophysiology* 47, 236–246. doi: 10.1111/j.1469-8986.2009.00928.x

Coffey, E. B., Herholz, S. C., Chepesiuk, A. M., Baillet, S., and Zatorre, R. J. (2016). Cortical contributions to the auditory frequency-following response revealed by MEG. *Nat. Commun.* 7:11070.

Coffey, E. B., Nicol, T., White-Schwoch, T., Chandrasekaran, B., Krizman, J., Skoe, E., et al. (2019). Evolving perspectives on the sources of the frequency-following response. *Nat. Commun.* 10:5036. doi: 10.1038/s41467-019-13003-w

Collet, L., and Duclaux, R. (1986). Auditory brainstem evoked responses and attention: contribution to a controversial subject. *Acta Otolaryngol. (Stockh.)* 101, 439–441.

Conway, C. M., Pisoni, D. B., and Kronenberger, W. G. (2009). The importance of sound for cognitive sequencing abilities. *Curr. Direct. Psychol. Sci.* 18, 275–279. doi: 10.1111/j.1467-8721.2009.01651.x

Cooke, M., Lecumberri, M. L. G., and Barker, J. (2008). The foreign language cocktail party problem: energetic and informational masking effects in non-native speech perception. *J. Acoust. Soc. Am.* 123, 414–27. doi: 10.1121/1.2804952

Costa, A., and Sebastián-Gallés, N. (2014). How does the bilingual experience sculpt the brain? *Nat. Rev. Neurosci.* 15, 336–345. doi: 10.1038/nrn3709

Crittenden, B. M., and Duncan, J. (2014). Task difficulty manipulation reveals multiple demand activity but no frontal lobe hierarchy. *Cereb. Cortex* 24, 532–540. doi: 10.1093/cercor/bhs333

D'Angiulli, A., Herdman, A., Stapells, D., and Hertzman, C. (2008). Children's event-related potentials of auditory selective attention vary with their socioeconomic status. *Neuropsychology* 22, 293–300. doi: 10.1037/0894-4105.22.3.293

Darwin, C. J. (1997). Auditory grouping. *Trends Cogn. Sci.* 1, 327–333.

Darwin, C. J., Brungart, D. S., and Simpson, B. D. (2003). Effects of fundamental frequency and vocal-tract length changes on attention to one of two simultaneous talkers. *J. Acoust. Soc. Am.* 114, 2913–2922. doi: 10.1121/1.1616924

de Abreu, P. M. E., Cruz-Santos, A., Tourinho, C. J., Martin, R., and Bialystok, E. (2012). Bilingualism enriches the poor enhanced cognitive control in low-income minority children. *Psychol. Sci.* 23, 1364–1371. doi: 10.1177/0956797612443836

de Bruin, A., Treccani, B., and Della Sala, S. (2015). Cognitive advantage in bilingualism an example of publication bias? *Psychol. Sci.* 26, 99–107. doi: 10.1177/0956797614557866

Dick, A. S., Garcia, N. L., Pruden, S. M., Thompson, W. K., Hawes, S. W., Sutherland, M. T., et al. (2019). No evidence for a bilingual executive function advantage in the ABCD study. *Nat. Hum. Behav.* 3, 692–701. doi: 10.1038/s41562-019-0609-3

Ding, N., and Simon, J. Z. (2014). Cortical entrainment to continuous speech: functional roles and interpretations. *Front. Hum. Neurosci.* 8:311. doi: 10.3389/fnhum.2014.00311

Edwards, E., Soltani, M., Deouell, L. Y., Berger, M. S., and Knight, R. T. (2005). High gamma activity in response to deviant auditory stimuli recorded directly from human cortex. *J. Neurophysiol.* 94, 4269–4280.

Figueras, B., Edwards, L., and Langdon, D. (2008). Executive function and language in deaf children. *J. Deaf Stud. Deaf Educ.* 13, 362–377. doi: 10.1093/deafed/enm067

Forte, A. E., Etard, O., and Reichenbach, T. (2017). The human auditory brainstem response to running speech reveals a subcortical mechanism for selective attention. *Elife* 6:e27203. doi: 10.7554/eLife.27203

Foy, J. G., and Mann, V. A. (2014). Bilingual children show advantages in nonverbal auditory executive function task. *Int. J. Biling.* 18, 717–729.

Fritz, J. B., Elhilali, M., David, S. V., and Shamma, S. A. (2007). Auditory attention - focusing the searchlight on sound. *Curr. Opin. Neurobiol.* 17, 437–455. doi: 10.1016/j.conb.2007.07.011

Galbraith, G. C., Bhuta, S. M., Choate, A. K., Kitahara, J. M., and Mullen, T. A. Jr. (1998). Brain stem frequency-following response to dichotic vowels during attention. *Neuroreport* 9, 1889–1893. doi: 10.1097/00001756-199806010-00041

García-Pentón, L., Fernández, A. P., Iturria-Medina, Y., Gillon-Dowens, M., and Carreiras, M. (2014). Anatomical connectivity changes in the bilingual brain. *Neuroimage* 84, 495–504.

Gathercole, S. E., and Baddeley, A. D. (2014). *Working Memory and Language*. East Sussex: Psychology Press.

Gold, B. T., Kim, C., Johnson, N. F., Kryscio, R. J., and Smith, C. D. (2013). Lifelong bilingualism maintains neural efficiency for cognitive control in aging. *J. Neurosci.* 33, 387–396. doi: 10.1523/JNEUROSCI.3837-12.2013

Golumbic, E. M. Z., Ding, N., Bickel, S., Lakatos, P., Schevon, C. A., McKhann, G. M., et al. (2013). Mechanisms underlying selective neuronal tracking of attended speech at a "cocktail party". *Neuron* 77, 980–991. doi: 10.1016/j.neuron.2012.12.037

Hagyard, J., Brimmell, J., Edwards, E. J., and Vaughan, R. S. (2021). Inhibitory control across athletic expertise and its relationship with sport performance. *J. Sport Exerc. Psychol.* 43, 14–27. doi: 10.1123/jsep.2020-0043

Hart, B., and Risley, T. R. (1995). *Meaningful Differences in the Everyday Experience of Young American Children*. Baltimore: Paul H Brookes Publishing.

Hartmann, T., and Weisz, N. (2019). Auditory cortical generators of the Frequency Following Response are modulated by intermodal attention. *Neuroimage* 203:116185. doi: 10.1016/j.neuroimage.2019.116185

Hoff, E. (2003). The specificity of environmental influence: socioeconomic status affects early vocabulary development via maternal speech. *Child Dev.* 74, 1368–1378. doi: 10.1111/1467-8624.00612

Hollingshead, A. (1975). *Four Factor Index of Social Status*. New Haven: Yale University.

Hopfinger, J., Buonocore, M., and Mangun, G. (2000). The neural mechanisms of top-down attentional control. *Nat. Neurosci.* 3, 284–291.

Ito, T., and Malmierca, M. S. (2018). "Neurons, connections, and microcircuits of the inferior colliculus," in *The Mammalian Auditory Pathways*, eds D. Oliver, N. Cant, R. Fay, and A. Popper (Cham: Springer), 127–167.

Jensen, O., Gelfand, J., Kounios, J., and Lisman, J. E. (2002). Oscillations in the alpha band (9–12 Hz) increase with memory load during retention in a short-term memory task. *Cereb. Cortex* 12, 877–882. doi: 10.1093/cercor/12.8.877

Jensen, O., and Hanslmyar, S. (2020). "The role of alpha oscillations for attention and working memory," in *The Cognitive Neurosciences*, eds D. Poeppel, G. R. Mangun, and M. S. Gazzaniga (Cambridge, MA: MIT Press), 323.

Jürgens, U. (1983). Afferent fibers to the cingular vocalization region in the squirrel monkey. *Exp. Neurol.* 80, 395–409. doi: 10.1016/0014-4886(83)90291-1

Kaushanskaya, M., Blumenfeld, H. K., and Marian, V. (2019). The Language Experience and Proficiency Questionnaire (LEAP-Q): Ten years later. *Biling. (Camb Engl)* 23, 945–950.

Khan, S., Edwards, L., and Langdon, D. (2005). The cognition and behaviour of children with cochlear implants, children with hearing aids and their hearing peers: a comparison. *Audiol. Neurootol.* 10, 117–126.

Kral, A., Kronenberger, W. G., Pisoni, D. B., and O'Donoghue, G. M. (2016). Neurocognitive factors in sensory restoration of early deafness: a connectome model. *Lancet Neurol.* 15, 610–621. doi: 10.1016/s1474-4422(16)00034-x

Krishnan, A., Gandour, J. T., Bidelman, G. M., and Swaminathan, J. (2009). Experience-dependent neural representation of dynamic pitch in the brainstem. *Neuroreport* 20, 408–413. doi: 10.1097/wnr.0b013e3283263000

Krishnan, A., Xu, Y., Gandour, J., and Cariani, P. (2005). Encoding of pitch in the human brainstem is sensitive to language experience. *Cogn. Brain Res.* 25, 161–168.

Krizman, J., Bradlow, A. R., Lam, S. S.-Y., and Kraus, N. (2017). How bilinguals listen in noise: linguistic and non-linguistic factors. *Biling. Lang. Cogn.* 20, 834–843. doi: 10.1017/s1366728916000444

Krizman, J., Marian, V., Shook, A., Skoe, E., and Kraus, N. (2012). Subcortical encoding of sound is enhanced in bilinguals and relates to executive function advantages. *Proc. Natl. Acad. Sci.* 109, 7877–7881. doi: 10.1073/pnas.1201575109

Krizman, J., Skoe, E., Marian, V., and Kraus, N. (2014). Bilingualism increases neural response consistency and attentional control: evidence for sensory and cognitive coupling. *Brain Lang.* 128, 34–40. doi: 10.1016/j.bandl.2013.11.006

Kroll, J. F., Bobb, S. C., Misra, M., and Guo, T. (2008). Language selection in bilingual speech: evidence for inhibitory processes. *Acta Psychol. (Amst.)* 128, 416–430. doi: 10.1016/j.actpsy.2008.02.001

Kronenberger, W. G., Xu, H., and Pisoni, D. B. (2020). Longitudinal development of executive functioning and spoken language skills in preschool-aged children with cochlear implants. *J. Speech Lang. Hear. Res.* 63, 1128–1147. doi: 10.1044/2019_JSLHR-19-00247

Langner, G., and Schreiner, C. E. (1988). Periodicity coding in the inferior colliculus of the cat. I. Neuronal mechanisms. *J. Neurophysiol.* 60, 1799–1822. doi: 10.1152/jn.1988.60.6.1799

Lecumberri, M. L. G., Cooke, M., and Cutler, A. (2011). Non-native speech perception in adverse conditions: a review. *Speech Commun.* 52, 864–886. doi: 10.1016/j.specom.2010.08.014

Liang-Fa, L., Alan, R. P., and Mark, N. W. (2006). Phase-locked responses to pure tones in the inferior colliculus. *J. Neurophysiol.* 95, 1926–35. doi: 10.1152/jn.00497.2005

Liegeois-Chauvel, C., Musolino, A., Badier, J., Marquis, P., and Chauvel, P. (1994). Evoked potentials recorded from the auditory cortex in man: evaluation and topography of the middle latency components. *Electroencephalogr. Clin. Neurophysiol.* 92, 204–214. doi: 10.1016/0168-5597(94)90064-7

Lind, R. R., Beck, M. M., Wikman, J., Malarski, K., Krustrup, P., Lundbye-Jensen, J., et al. (2019). Acute high-intensity football games can improve children's inhibitory control and neurophysiological measures of attention. *Scand. J. Med. Sci. Sports* 29, 1546–1562. doi: 10.1111/sms.13485

Lucks Mendel, L., and Widner, H. (2016). Speech perception in noise for bilingual listeners with normal hearing. *Int. J. Audiol.* 55, 126–134. doi: 10.3109/14992027.2015.1061710

Malmierca, M. S. (2015). "Anatomy and physiology of the mammalian auditory system," in *Encyclopedia of Computational Neuroscience*, eds D. Jaeger and R. Jung, (New York: Springer), 155–186. doi: 10.1007/978-1-4614-6675-8_286

Malmierca, M. S., Cristaudo, S., érez-González, D. P., and Covey, E. (2009). Stimulus-specific adaptation in the inferior colliculus of the anesthetized rat. *J. Neurosci.* 29, 5483–5493. doi: 10.1523/jneurosci.4153-08.2009

Malmierca, M. S., Nino-Aguillon, B. E., Nieto-Diego, J., Porteros, Á, Pérez-González, D., and Escera, C. (2019). Pattern-sensitive neurons reveal encoding of complex auditory regularities in the rat inferior colliculus. *Neuroimage* 184, 889–900. doi: 10.1016/j.neuroimage.2018.10.012

Malmierca, M. S., and Ryugo, D. K. (2011). "Descending connections of auditory cortex to the midbrain and brain stem," in *The Auditory Cortex*, eds J. Winer, and C. Schreiner, (New York: Springer), 189–208. doi: 10.1007/978-1-4419-0074-6_9

Marian, V., Blumenfeld, H. K., and Kaushanskaya, M. (2007). The language experience and proficiency questionnaire (LEAP-Q): assessing language profiles in bilinguals and multilinguals. *J. Speech Lang. Hear. Res.* 50, 940–67. doi: 10.1044/1092-4388(2007/067)

Marian, V., Lam, T. Q., Hayakawa, S., and Dhar, S. (2018). Top-down cognitive and linguistic influences on the suppression of spontaneous otoacoustic emissions. *Front. Neurosci.* 12:378. doi: 10.3389/fnins.2018.00378

Marian, V., and Spivey, M. (2003). Bilingual and monolingual processing of competing lexical items. *Appl. Psycholing.* 24, 173–193. doi: 10.1017/s0142716403000092

Mattys, S. L., Davis, M. H., Bradlow, A. R., and Scott, S. K. (2012). Speech recognition in adverse conditions: a review. *Lang. Cognit. Process.* 27, 953–978. doi: 10.1080/01690965.2012.705006

Mayo, L. H., Florentine, M., and Buus, S. (1997). Age of second-language acquisition and perception of speech in noise. *J. Speech Lang. Hear. Res.* 40, 686–93. doi: 10.1044/jslhr.4003.686

Mechelli, A., Crinion, J. T., Noppeney, U., O'Doherty, J., Ashburner, J., Frackowiak, R. S., et al. (2004). Neurolinguistics: structural plasticity in the bilingual brain. *Nature* 431, 757–757.

Mesgarani, N., and Chang, E. F. (2012). Selective cortical representation of attended speaker in multi-talker speech perception. *Nature* 485, 233–236.

Mitchell, T. V., and Quittner, A. L. (1996). Multimethod study of attention and behavior problems in hearing-impaired children. *J. Clin. Child Psychol.* 25, 83–96.

Moreno, S., Wodniecka, Z., Tays, W., Alain, C., and Bialystok, E. (2014). Inhibitory control in bilinguals and musicians: event related potential (ERP) evidence for experience-specific effects. *PLoS One* 9:e94169. doi: 10.1371/journal.pone.0094169

Morini, G. (2020). Examining bilingual speech perception and the role of background noise. *J. Acoust. Soc. Am.* 148:2623.

Neill, W., Valdes, L., and Terry, K. (1995). "Selective attention and the inhibitory control of cognition," in *Interference and Inhibition in Cognition*. eds F. N. Dempster and C. J. Brainerd, (San Diego: Academic Press), 207–261.

Olguin, A., Cekic, M., Bekinschtein, T. A., Katsos, N., and Bozic, M. (2019). Bilingualism and language similarity modify the neural mechanisms of selective attention. *Sci. Rep.* 9:8204.

Osterhammel, P., Shallop, J., and Terkildsen, K. (1985). The effect of sleep on the auditory brainstem response (ABR) and the middle latency response (MLR). *Scand. Audiol.* 14, 47–50. doi: 10.3109/01050398509045921

Paap, K. R., Johnson, H. A., and Sawi, O. (2015). Bilingual advantages in executive functioning either do not exist or are restricted to very specific and undetermined circumstances. *Cortex* 69, 265–278. doi: 10.1016/j.cortex.2015.04.014

Parras, G. G., Nieto-Diego, J., Carbajal, G. V., Valdés-Baizabal, C., Escera, C., and Malmierca, M. S. (2017). Neurons along the auditory pathway exhibit a hierarchical organization of prediction error. *Nat. Commun.* 8:2148. doi: 10.1038/s41467-017-02038-6

Pichora-Fuller, M. K., Kramer, S. E., Eckert, M. A., Edwards, B., Hornsby, B. W., Humes, L. E., et al. (2016). Hearing impairment and cognitive energy: the framework for understanding effortful listening (FUEL). *Ear Hear.* 37, 5S–27S.

Price, C. N., and Bidelman, G. M. (2021). Attention reinforces human corticofugal system to aid speech perception in noise. *Neuroimage* 235:118014. doi: 10.1016/j.neuroimage.2021.118014

Raizada, R. D., and Poldrack, R. A. (2007). Challenge-driven attention: interacting frontal and brainstem systems. *Front. Hum. Neurosci.* 1:3. doi: 10.3389/neuro.09.003.2007

Ressel, V., Pallier, C., Ventura-Campos, N., Díaz, B., Roessler, A., Ávila, C., et al. (2012). An effect of bilingualism on the auditory cortex. *J. Neurosci.* 32, 16597–16601. doi: 10.1523/jneurosci.1996-12.2012

Salamy, A., and McKean, C. M. (1977). Habituation and dishabituation of cortical and brainstem evoked potentials. *Int. J. Neurosci.* 7, 175–182. doi: 10.3109/00207457709147665

Shi, L.-F. (2010). Perception of acoustically degraded sentences in bilingual listeners who differ in age of English acquisition. *J. Speech Lang. Hear. Res.* 53, 821–35. doi: 10.1044/1092-4388(2010/09-0081)

Shi, L.-F. (2012). Contribution of linguistic variables to bilingual listeners' perception of degraded english sentences. *J. Speech Lang. Hear. Res.* 55, 219–34. doi: 10.1044/1092-4388(2011/10-0240)

Skoe, E. (2019). Turn up the volume: speech perception in noise for bilingual listeners. *J. Acoust. Soc. Am.* 145, 1820–1820.

Skoe, E., Burakiewicz, E., Figueiredo, M., and Hardin, M. (2017). Basic neural processing of sound in adults is influenced by bilingual experience. *Neuroscience* 349, 278–290. doi: 10.1016/j.neuroscience.2017.02.049

Skoe, E., and Karayanidi, K. (2019). Bilingualism and speech understanding in noise: auditory and linguistic factors. *J. Am. Acad. Audiol.* 30, 115–130.

Slater, J., Ashley, R., Tierney, A., and Kraus, N. (2018). Got rhythm? Better inhibitory control is linked with more consistent drumming and enhanced neural tracking of the musical beat in adult percussionists and nonpercussionists. *J. Cogn. Neurosci.* 30, 14–24. doi: 10.1162/jocn_a_01189

Slee, S. J., and David, S. V. (2015). Rapid task-related plasticity of spectrotemporal receptive fields in the auditory midbrain. *J. Neurosci.* 35, 13090–13102. doi: 10.1523/JNEUROSCI.1671-15.2015

Song, J. H., Skoe, E., Banai, K., and Kraus, N. (2010). Perception of speech in noise: neural correlates. *J. Cogn. Neurosci.* 23, 2268–2279.

Song, J. H., Skoe, E., Wong, P. C. M., and Kraus, N. (2008). Plasticity in the adult human auditory brainstem following short-term linguistic training. *J. Cogn. Neurosci.* 20, 1892–1902.

Spivey, M., and Marian, V. (1999). Cross talk between native and second languages: partial activation of an irrelevant lexicon. *Psychol. Sci.* 10, 281–284. doi: 10.1111/1467-9280.00151

Strori, D., Bradlow, A. R., and Souza, P. E. (2020). Recognition of foreign-accented speech in noise: the interplay between talker intelligibility and linguistic structure. *J. Acoust. Soc. Am.* 147, 3765–3782. doi: 10.1121/10.0001194

Stuart, A., Zhang, J., and Swink, S. (2010). Reception thresholds for sentences in quiet and noise for monolingual english and bilingual Mandarin-English listeners. *J. Am. Acad. Audiol.* 21, 239–248. doi: 10.3766/jaaa.21.4.3

Thierry, G., and Wu, Y. J. (2007). Brain potentials reveal unconscious translation during foreign-language comprehension. *Proc. Natl. Acad. Sci. U. S. A.* 104, 12530–12535. doi: 10.1073/pnas.0609927104

Tun, P. A., O'Kane, G., and Wingfield, A. (2002). Distraction by competing speech in young and older adult listeners. *Psychol. Aging* 17, 453–67.

Van Heuven, W. J., Schriefers, H., Dijkstra, T., and Hagoort, P. (2008). Language conflict in the bilingual brain. *Cereb. Cortex* 18, 2706–2716.

Varghese, L., Bharadwaj, H. M., and Shinn-Cunningham, B. G. (2015). Evidence against attentional state modulating scalp-recorded auditory brainstem steady-state responses. *Brain Res.* 1626, 146–164. doi: 10.1016/j.brainres.2015.06.038

Wechsler, D. (1999). *Wechsler Abbreviated Scale of Intelligence (WASI)*. San Antonio: Harcourt Assessment.

Weissman, D. H., Roberts, K. C., Visscher, K. M., and Woldorff, M. G. (2006). The neural bases of momentary lapses in attention. *Nat. Neurosci.* 9, 971–978. doi: 10.1038/nn1727

White-Schwoch, T., Anderson, S., Krizman, J., Nicol, T., and Kraus, N. (2019). Case studies in neuroscience: subcortical origins of the frequency-following response. *J. Neurophysiol.* 122, 844–848. doi: 10.1152/jn.00112.2019

Wolfe, C. D., and Bell, M. A. (2004). Working memory and inhibitory control in early childhood: contributions from physiology, temperament, and language. *Dev. Psychobiol.* 44, 68–83. doi: 10.1002/dev.10152

Wu, C., Weissman, D., Roberts, K., and Woldorff, M. (2007). The neural circuitry underlying the executive control of auditory spatial attention. *Brain Res.* 1134, 187–98. doi: 10.1016/j.brainres.2006.11.088

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Krizman, Tierney, Nicol and Kraus. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 20 December 2021
doi: 10.3389/fnins.2021.747303

Rapid Enhancement of Subcortical Neural Responses to Sine-Wave Speech

Fan-Yin Cheng†, Can Xu†, Lisa Gold† and Spencer Smith†*

Department of Speech, Language, and Hearing Sciences, University of Texas at Austin, Austin, TX, United States

The efferent auditory nervous system may be a potent force in shaping how the brain responds to behaviorally significant sounds. Previous human experiments using the frequency following response (FFR) have shown efferent-induced modulation of subcortical auditory function online and over short- and long-term time scales; however, a contemporary understanding of FFR generation presents new questions about whether previous effects were constrained solely to the auditory subcortex. The present experiment used sine-wave speech (SWS), an acoustically-sparse stimulus in which dynamic pure tones represent speech formant contours, to evoke FFR_{SWS}. Due to the higher stimulus frequencies used in SWS, this approach biased neural responses toward brainstem generators and allowed for three stimuli (/bɔ/, /bu/, and /bo/) to be used to evoke FFR_{SWS} *before* and *after* listeners in a training group were made aware that they were hearing a degraded speech stimulus. All SWS stimuli were rapidly perceived as speech when presented with a SWS carrier phrase, and average token identification reached ceiling performance during a perceptual training phase. Compared to a control group which remained naïve throughout the experiment, training group FFR_{SWS} amplitudes were enhanced post-training for each stimulus. Further, linear support vector machine classification of training group FFR_{SWS} significantly improved post-training compared to the control group, indicating that training-induced neural enhancements were sufficient to bolster machine learning classification accuracy. These results suggest that the efferent auditory system may rapidly modulate auditory brainstem representation of sounds depending on their context and perception as non-speech or speech.

Keywords: frequency following response (FFR), efferent, top-down, sine-wave speech perception, auditory learning

OPEN ACCESS

Edited by:
Erika Skoe,
University of Connecticut, United States

Reviewed by:
Zilong Xie,
University of Kansas Medical Center, United States
Hilmi Dajani,
University of Ottawa, Canada

***Correspondence:**
Spencer Smith
spencer.smith@austin.utexas.edu

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Auditory Cognitive Neuroscience, a section of the journal Frontiers in Neuroscience

Received: *26 July 2021*
Accepted: *02 December 2021*
Published: *20 December 2021*

Citation:
Cheng F-Y, Xu C, Gold L and Smith S (2021) Rapid Enhancement of Subcortical Neural Responses to Sine-Wave Speech. Front. Neurosci. 15:747303. doi: 10.3389/fnins.2021.747303

INTRODUCTION

The mammalian auditory system contains extensive efferent innervation descending from the cortex to subcortex and inner ear (Winer, 2005). Numerous animal modeling studies suggest that these projections facilitate neuroplastic functional changes on multiple time scales and at multiple levels of the subcortical auditory system. For example, *online* modulation of auditory function has been observed at the level of the cochlea (Xiao and Suga, 2002; May et al., 2004; Dragicevic et al., 2015; Terreros and Delano, 2015; Delano and Elgoyhen, 2016; Lauer et al., 2021), cochlear nucleus

(Hernandez-Peon et al., 1956), and inferior colliculus (Slee and David, 2015; Shaheen et al., 2021). *Short-* and *long-term* training also alters physiologic function in the same structures (Gao and Suga, 1998, 2000; Yan and Suga, 1998, 1999; Suga et al., 2000, 2002; Ji et al., 2001; Ma and Suga, 2001; Yan et al., 2005; Malmierca et al., 2009). Inversely, obliterating or temporarily silencing corticofugal efferent connections disrupts online modulation and short- and long-term training effects measured subcortically (e.g., Bajo et al., 2010; León et al., 2012). Together, these studies suggest that efferent activity is a potent force in shaping how the nervous system responds to behaviorally significant sounds, even at the earliest stages of auditory processing.

Efferent-induced changes in human subcortical auditory function have, by necessity, almost exclusively been assessed through non-invasive objective measurements. Some reports have demonstrated that otoacoustic emissions (i.e., proxy measures of outer hair cell function) are modulated online by attention (Wittekindt et al., 2014; Smith and Cone, 2015; Hernandez-Perez et al., 2021) or through short- or long-term training (Perrot et al., 2006; de Boer and Thornton, 2008; Bidelman et al., 2014, 2016, 2017). Other reports using similar methodologies have failed to replicate these findings (Stuart and Butler, 2012; Francis et al., 2018; Jedrzejczak et al., 2020). A variety of electrophysiologic measures has been used to study online or training-based neuroplastic functional changes in the human auditory subcortex including the auditory brainstem response (ABR) and frequency following response (FFR). As a general principle, the "classic" ABR does not appear to be altered by attention (Picton and Hillyard, 1974; Woldorff et al., 1987; Connolly et al., 1989; Gregory et al., 1989; Hackley et al., 1990), whereas the FFR literature presents a less cohesive narrative. Seminal work by Galbraith and Arroyo (1993) and Galbraith et al. (1995, 1998, 2003) suggested that FFRs to simple (e.g., tonal) and complex (e.g., dichotic speech) stimuli were modulated during auditory or visual attention. While some researchers have replicated these findings (e.g., Hairston et al., 2013; Lehmann and Schönwiesner, 2014), others have failed to observe attention effects and have questioned whether previous results were influenced by task-based differences in FFR residual noise (Ruggles et al., 2012; Varghese et al., 2015). More recent studies demonstrate FFR enhancements during active listening to ecologically valid continuous speech (Forte et al., 2017; Etard et al., 2019; Saiz-Alía et al., 2019).

A larger body of FFR literature supports the supposition that short- and long-term training induce neuroplastic changes in the auditory subcortex over time. Studies in which listeners were trained to discriminate stimuli by focusing on a specific sound feature (e.g., global pitch or dynamic pitch contours) have reported enhancement of the neural representation of the trained feature (e.g., Russo et al., 2005; Song et al., 2008, 2012; Carcagno and Plack, 2011; Chandrasekaran et al., 2012; Skoe et al., 2014). These changes were noted after multiple hours or days of training; however, additional studies have reported rapid FFR modulation occurring within minutes of training onset (e.g., Skoe and Kraus, 2010; Skoe et al., 2013). Similar and more robust enhancements are observed in musicians (Wong et al., 2007; Bidelman and Krishnan, 2009) and tonal language speakers (Krishnan et al., 2005; Swaminathan et al., 2008; Krishnan and Gandour, 2009) who, by virtue of their lived experiences, have undergone a form of long-term auditory training (see Kraus and Chandrasekaran, 2010; Strait and Kraus, 2014; Kraus and White-Schwoch, 2015 for reviews).

The majority of FFR studies examining online or training-related changes in neural function have focused on neural representation of the speech envelope and its harmonics (FFR_{ENV}). Recent evidence suggests that although the FFR_{ENV} arises primarily from the auditory subcortex (Chandrasekaran and Kraus, 2010; Bidelman, 2015; Bidelman and Powers, 2018; Bidelman et al., 2018a), cortical contributions may also be present, particularly for stimuli with fundamental frequencies $< \sim$150 Hz (Coffey et al., 2016, 2019). This new understanding of FFR_{ENV} origins presents the possibility that neuroplastic changes observed in some previous studies may not be constrained to the subcortex. One way to ensure that measured neural responses are biased exclusively toward subcortical generators is to use stimuli comprised of behaviorally-significant higher frequency ($>$200 Hz) speech content, as more caudal generators begin to dominate the FFR with increasing stimulus frequencies (Gardi et al., 1979; Galbraith et al., 2001; Tichko and Skoe, 2017).

Sine-wave speech (SWS) is an acoustically manipulated form of speech in which formant trajectories are represented by time-variant sine waves, and the remainder of the acoustic signal is discarded (Remez et al., 1981). It can therefore be conceptualized as speech "fine structure" that has been spectrally reduced to two or three dynamic frequency components. The range of average first (F1) and second (F2) formant frequencies in adult American English speakers is $\sim$300–775 and $\sim$900–2,700 Hz, respectively (Lindblom, 1990). Because the upper frequency limit of the FFR is approximately $\sim$1,200–1,300 Hz (Bidelman and Powers, 2018), much of the F1 and F2 formant space may be captured by FFRs evoked by SWS (FFR_{SWS}). A critical advantage of SWS is that naïve listeners do not hear it as speech (Remez et al., 1981; Barker and Cooke, 1999; Möttönen et al., 2006); however, with minimal instruction and/or training, listeners achieve a high level of SWS comprehension. Consequently, it is possible to use identical speech-like stimuli to evoke FFR_{SWS} pre- and post-engagement of the auditory efferent system through online or brief short-term training activities. While the neural networks involved in this top-down process are not fully understood, recent reports examining cortical responses to SWS (or vocoded speech) indicate that activity from different brain networks is involved based on whether the signals are perceived as speech or non-speech (Davis and Johnsrude, 2003; Eisner et al., 2010; Hervais-Adelman et al., 2012; Khoshkhoo et al., 2018). Specifically, SWS is represented in the auditory cortex based on "bottom-up" acoustic features in naïve listeners. When listeners undergo a perceptual shift and begin to understand these degraded signals as speech, left inferior frontal cortex activity increases significantly while auditory cortex activity remains stable (Khoshkhoo et al., 2018). Given the observations that active listening sequentially modulates neural tuning in the same cortical networks in a top-down fashion (e.g., Atiani et al., 2014), it is possible that these modulatory effects continue into the auditory brainstem *via* the efferent system (Bidelman et al., 2019; Price and Bidelman, 2021).

In the present experiment, three SWS tokens, differing mainly in their F1 contours, were used to evoke FFR_{SWS} before and after a brief auditory training paradigm in which listeners were informed that they were listening to degraded speech and were asked to classify each token. These results were compared to FFR_{SWS} measured from a control group, which did not undergo training. FFR_{SWS} were confirmed to be of neural origin with latencies suggesting brainstem generators and high stimulus-to-response cross-correlations. In the test group, all SWS stimuli were rapidly perceived as speech when presented with a SWS carrier phrase in a brief training phase, and average token identification reached ceiling performance within 25 training trials or less per stimulus. FFR_{SWS} amplitudes in the test group were enhanced post-training for each stimulus compared to the control group. Further, linear support vector machine classification of FFR_{SWS} significantly improved post-training in the test group compared to controls, indicating that training-induced neural enhancements were sufficient to bolster machine learning classification accuracy. These results suggest that the efferent auditory system may rapidly modulate auditory brainstem representation of sounds depending on their context and perception as non-speech or speech.

MATERIALS AND METHODS

Participants

This study was approved by the University of Texas at Austin Institutional Review Board. Eighteen adults (mean age = 22.2 years) with no history of audiologic or neurologic injury were enrolled. Half of the participants were placed in a training group and the other half served as untrained controls. Participants had normal hearing (≤25 dB HL) from 250 to 8,000 Hz bilaterally. Each participant provided written consent and completed 3 h of testing for which they were compensated.

Sine Wave Speech Stimuli

Three naturally produced CV speech tokens, /bɔ/, /bu/, and /bo/, were recorded (44,100 Hz sampling rate) from an adult male speaker with a Standard American English accent. The speaker was told to maintain constant voice pitch across all recordings. Each CV token was 335 ms in duration, and cosine squared ramps were applied to the last 50 ms of each stimulus to equate and smooth offsets across stimuli. The natural CV tokens were then converted to SWS in Praat software (Boersma, 2009) using the approach developed by Darwin (2003). This approach uses linear predictive coding analysis to identify formant center frequencies and amplitudes within a sliding window over the stimulus. The formants are then replaced with time-varying sinusoids, and all other speech content is discarded (**Figure 1**). Only the first two formants from the original stimuli were kept, as FFRs were unlikely to be evoked by higher frequency formants. A carrier phrase ("The word is ____.") that was only used in the brief training phase for the training group (described below) was also converted to SWS in the same manner described above. All SWS stimuli were RMS normalized to ensure equal presentation level.

The three vowels in the CV stimuli were selected for multiple reasons related to their relative positions in the F1/F2 formant space. First, phase-locking in the auditory nervous system becomes poorer as stimulus frequency increases. Consequently, stimuli comprised of lower frequencies generate more robust FFR responses (e.g., Bidelman and Powers, 2018). The vowels /ɔ/, /u/, and /o/ have the lowest possible F1 and F2 frequencies in American English and are therefore the most ideal SWS candidates for evoking robust $FFRs_{WS}$. Second, the vowels primarily differ in their F1 contours, whereas the F2 contours are less disparate. The range in F1 frequencies for the three vowels was ∼300–675 Hz, whereas the range in F2 was ∼1,100–1,300 Hz. Because CV differences were most pronounced in their F1 frequencies, we anticipated that listeners in the training group would primarily focus on this feature to successfully complete the auditory training task and that neural enhancement related to the brief training period would be apparent at the F1 frequency (described below). Third, the total range of F1 and F2 stimulus frequencies (∼300–1,300 Hz) biases the FFR to reflect more caudal subcortical generators (e.g., Galbraith et al., 2001; Bidelman, 2018).

Procedure

Training Group

All experimental procedures occurred in a double-walled sound booth with participants seated in a reclining chair. Auditory stimuli were presented diotically through electromagnetically shielded ER-3 insert earphones (Etymotic Research, Elk Grove Village, IL), and visual prompts (used only in the training phase) were presented through a Dell PC monitor. Experiment stimuli were programmed and controlled *via* Neuroscan's GenTask module (Compumedics Neuroscan, Charlotte, NC). The experiment began with a *pre-training phase* in which FFR_{SWS} were evoked by /bɔ/, /bu/, and /bo/ SWS tokens presented in random order. Half of the stimulus presentations were in one polarity ("Polarity A") and half were in the opposite polarity ("Polarity B"). Each stimulus was presented in each polarity 1,000 times for a total of 6,000 sweeps. The intertrial interval between stimuli was 600 ms. During the pre-training phase, each participant was asked to remain still while quietly watching a subtitled movie or show of his or her choosing.

The *training phase* of the experiment began upon conclusion of the pre-training phase and after a brief break. Participants were notified that the stimuli they were hearing in the previous block were modified speech signals and that the training phase would require them to learn and identify the speech signals using a response keypad (Compumedics Neuroscan, Charlotte, NC). No additional instruction was given. At the beginning of each training trial, participants heard the SWS carrier phrase *"The word is _____."* with one of the three SWS tokens randomly presented as the target word. Simultaneously to the auditory presentation of the carrier phrase and target word, the participants saw a visual prompt on a monitor located directly in front of them and outside of the sound booth. The prompt depicted a visual representation of the carrier phrase with a blank in the target word space, exactly as written in

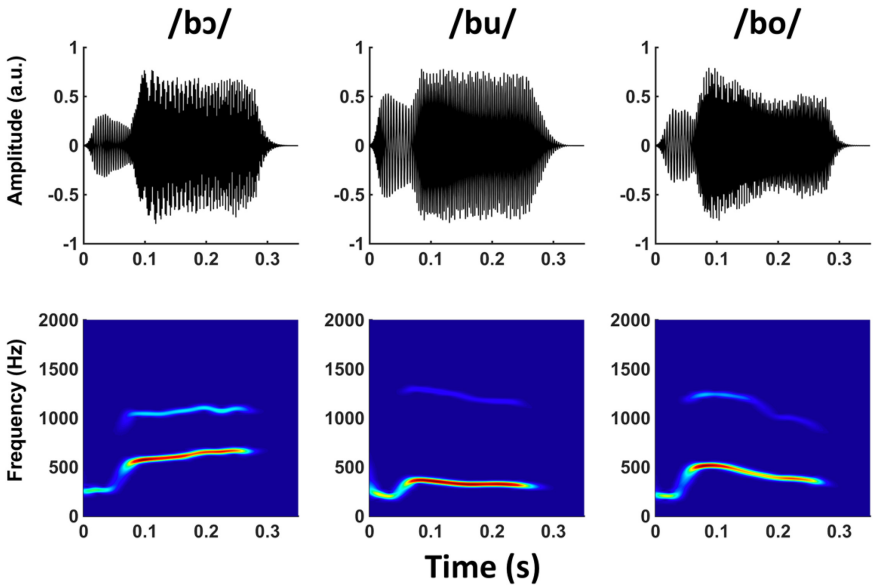

FIGURE 1 | Waveforms and spectrograms of /bɔ/, /bu/, and /bo/ SWS stimuli.

the italicized quote above. The total duration of the carrier phrase and target word was 1,500 ms; an additional 500 ms of silence was appended to the end of each carrier and target presentation to encourage participants to remain still prior to pressing the response keypad to submit their answer following the next prompt. Participants then saw a slide on the monitor with possible target words written non-phonetically as "bah," "boo," or "bow." The participant indicated which SWS word was heard by pressing one of three buttons on the response keypad, which was then followed by a 600 ms intertrial interval. Visual feedback ("Correct" or "Incorrect") was then given to participants, which was followed by another 600 ms interval before the onset of the next trial. This procedure was repeated 25 times per stimulus for a total of 75 training trials. Relatively few training trials were chosen based on previous reports that SWS becomes rapidly intelligible with very little training (Remez et al., 1981; Möttönen et al., 2006).

A *testing phase* followed the training phase. The main purpose of the testing phase was to ensure that participants retained SWS identification accuracy in the absence of the carrier phrase, which provided additional "samples" of the speaker's formant structure. In the testing phase, each trial began with the random presentation of a SWS target token. After a 600 ms pause, participants were invited to indicate their responses on a keypad, using the same slide described above with written target words as a reference. Participants had 900 ms from the onset to indicate their responses. Feedback was not provided in the test phase. Following training and testing phases, a *post-training phase*, parametrically identical to the pre-training phase, was conducted.

Control Group

The control group underwent passive FFR_{SWS} measurements that were identical to pre-training and post-training measurements in the training group. In place of the SWS training and testing phases, the control group was asked to watch an unrelated captioned television show and answer comprehension questions related to its content. While control group participants watched the captioned television show, they were exposed to the same carrier sentences as the test-group; however, they were never told that they were hearing modified speech at any point of the experiment. None of the control participants perceived the SWS stimuli to be speech according to a post-experiment survey. The purpose of including the control group in this study was to determine if pre- and post-training FFR_{SWS} enhancements in the test group were simply related to *exposure* to the SWS stimuli during the recording session and not due to efferent modulation following a perceptual shift from non-speech to speech perception. Note that, for simplicity, we refer to the first and second passive FFR_{SWS} measurements for test and control groups as "pre- and post-training" measurements throughout the manuscript, even though the control group did not undergo auditory training.

EEG Acquisition and Pre-processing

Electrophysiologic responses were obtained with a Neuroscan SynAmps2 system (Compumedics Neuroscan, Charlotte, NC). Responses were recorded at a 5,000 Hz sampling rate *via* a single-channel bipolar montage, Fpz (+), C7 vertebra (−), forehead (GND), and amplified by a factor of 100,000. Continuous data were exported from Curry 8 software, and further analyses were performed offline in MATLAB (The MathWorks, Natick, MA). All continuous data were first bandpass filtered from 100 to 2,400 Hz. For pre- and post-training FFR_{SWS}, continuous responses were epoched from −50 to 550 ms (re: SWS token onset), and single-trial responses were grouped by stimulus type. Responses were corrected for insert earphone delays by

subtracting 1 ms from the epoched data. Epochs were detrended, artifact rejected at ± 50 μV, and baseline corrected. Remaining sweeps were used to create grand average FFR_{SWS} for each stimulus such that individual polarities (A and B) as well as "added" [(A+B)/2] and subtracted [(A−B)/2] waveforms could be independently evaluated. Individual polarities and subtracted waveforms were used in a cross-correlation analysis (described below) to verify that FFR_{SWS} were neural in origin. Added polarity responses are generally used to accentuate neural representation of the envelope (Aiken and Picton, 2008). Because the stimuli in the present study did not have envelopes, added polarity waveforms were evaluated mainly as a quality control measure to ensure that FFR_{SWS} were not obliterated (which indicates that the measured responses are stimulus artifact or cochlear microphonic). In some cases, low amplitude waveforms containing energy at F1*2 were observed in the added polarity. This likely occurs because phase locked neural responses evoked by one stimulus polarity are temporally shifted by a half-cycle relative to the opposite polarity due to half-wave rectification (see Aiken and Picton, 2008; Lichtenhan et al., 2013). Adding these responses together can produce a doubling of the stimulus frequency and provides additional evidence that the measured responses are from *neural* generators.

ANALYSES

Test Group Training- and Testing-Phase Response Accuracy and Reaction Time

Training group response accuracy and reaction time were evaluated using behavioral data from training and test phases, respectively, as both measures are indicative of auditory training effects (e.g., Ritter et al., 1972; Song et al., 2008). Response accuracy, defined binarily on each trial as "correct" or "incorrect," was analyzed using mixed effects logistic regression with trial number and stimulus type as independent variables. Reaction time, defined as the post-stimulus onset time (re: to SWS target) at which respondents pressed the response keypad to indicate their choice, was evaluated using multiple linear regression with trial number and stimulus type as independent variables.

Stimulus-to-Response Cross-Correlation

Electromagnetic stimulus artifact, cochlear microphonic, and FFR waveforms can all mimic periodic characteristics of the input stimulus. A common method used to evaluate whether measured electrophysiologic responses are from neural generators or non-neural contaminants is to perform a cross-correlation between the stimulus and response. In this procedure, correlations between stimulus and FFR waveforms are calculated as the FFR waveform is temporally shifted relative to the stimulus waveform on a point-by-point basis (Skoe and Kraus, 2010). The time lag that produces the largest correlation coefficient is an estimated delay between stimulus and response. Responses generated by the auditory nerve and brainstem are expected to have a delay of ∼3–10 ms, depending on the electrode montage, stimulus frequency, and interaction between multiple neural generators as they reach scalp electrodes (e.g., Galbraith et al., 2001; Tichko and Skoe, 2017; Bidelman, 2018). In contrast, cochlear microphonic (arising from hair cell alternating currents primarily in the basal tail of the basilar membrane traveling wave; see Eggermont, 2017 for review) and stimulus artifact have short delays of ∼0–1 ms (Gardi et al., 1979). Stimulus-to-response cross-correlations were calculated for individual polarities (A and B) and subtracted waveforms evoked by each SWS stimulus in the pre- and post-training phases for test and control groups. SWS stimuli were first down-sampled from 44,100 to 5,000 Hz to match the FFR_{SWS} sampling rate, resulting in 0.2 ms precision in delay estimates. The maximum possible time delay producing the largest correlation coefficient was constrained between ±20 ms. Responses for which the estimated delays were within 3–10 ms were considered to be of neural origin. These responses were kept for further analysis. Cross-correlation coefficients, which are constrained between −1 and 1 and are non-normally distributed, were transformed to Fisher *z*-values (Cohen et al., 2013). A three-way multiple analysis of variance (MANOVA) with repeated measures was conducted to assess the impacts of group (test vs. control). training status (pre- vs. post-training), and stimuli (/bɔ/, /bu/, and /bo/) on participants' FFR_{SWS} latencies and z-transformed cross-correlation coefficients.

Fourier Analyzer

In contrast to a Fourier transform, which is commonly used to analyze steady-state stimuli/responses, a Fourier analyzer (FA) provides a better estimate of response amplitudes at frequencies of interest for signals with time-varying spectra (Aiken and Picton, 2006). Because FFRs are expected to follow dynamic frequency changes of a stimulus over time, the FA uses the stimulus frequency trajectory as a "reference" to detect FFR spectral amplitudes at frequencies along this trajectory by integration (Aiken and Picton, 2006). The stimuli used in the present study have non-stationary F1 and F2. Therefore, an FA was implemented to calculate response amplitudes in frequency bins corresponding to F1 and F2 trajectories to determine the strength of neural phase locking to each simulated formant.

We used a similar approach to implement FA as described by Aiken and Picton (2006) and Choi et al. (2013). First, stimulus reference tracks following F1 and F2, respectively, were created by exporting only F1 or F2 SWS sine-waves from Praat. Complex representations of F1 and F2 stimuli were obtained by Hilbert transform, and the instantaneous phase was calculated by finding the angle of the output of the Hilbert transform. F1 and F2 instantaneous frequencies were then calculated as the derivatives of the unwrapped phases at each time point. Since calculating the derivatives in this manner is equivalent to applying a high-pass filter, it introduces sharp perturbations in the resulting frequency tracks. Consequently, we smoothed the obtained instantaneous frequencies across time by applying a 50-point boxcar moving average 3 times. Reference complex sinusoids were then created for F1 and F2 frequency tracks using the instantaneous phase angles for each stimulus.

As mentioned above, FFRs demonstrate a characteristic delay between stimulus and response of ∼3–10 ms due to neural conduction time. FFR_{SWS} waveforms were shifted by –

6 ms based on pilot data testing to correct for neural delays and ensure that reference tracks were, on average, temporally aligned with the FFR_{SWS} waveforms prior to integration (Purcell et al., 2004; Aiken and Picton, 2006). Reference tracks and FFR_{SWS} waveforms were then integrated by multiplying the two waveforms over time (Choi et al., 2013) and computing the mean of the obtained complex numbers. The absolute value and the angle of the mean were then calculated as the FFR amplitude and the phase over the duration of the response (50–335 ms), respectively.

In order to determine whether FFR_{SWS} amplitudes at F1 and F2 were above background noise levels, 10 adjacent frequency tracks were also created to measure response amplitudes at non-stimulus frequencies. Five noise tracks above and five below each F1 and F2 track were obtained by adding or subtracting a fixed number of cycles per second. Noise tracks began at F1 $\pm$ 5 and F2 $\pm$ 5 Hz, respectively, and increased or decreased in 1 Hz steps. The same FA procedures as above were then used to estimate noise levels in the 10 adjacent non-stimulus frequency bins. F1 and F2 responses were deemed "present" if their amplitudes exceeded the noise floor averaged across 10 adjacent frequency bins. This approach is more lenient than other statistically based methods for determining response presence/absence (e.g., *F*-tests or Hotelling's T^2-tests; see Picton et al., 2003 for review). However, because the primary focus of this study was to evaluate potential *enhancement* of FFR_{SWS} following perceptual shifts, we did not want to remove participants who had low baseline FFR_{SWS}.

Machine Learning Classification of FFR_{SWS}

Previous experiments have used machine learning algorithms to assess whether the information contained in FFRs is sufficient to decode the stimulus classes that evoked them (Sadeghian et al., 2015; Holdgraf et al., 2017; Llanos et al., 2017; Yi et al., 2017; Xie et al., 2018, 2019). Under this approach, FFR classification performance (i.e., the accuracy with which FFRs are correctly classified by the machine learning algorithm) serves as an objective measure of stimulus discrimination. Importantly, FFR classification accuracy can be compared between levels of an independent variable (e.g., training or attention conditions) to determine how these factors impact classification performance (e.g., Xie et al., 2018). The rationale is that, if attention or training modulate neural function as captured by the FFR, the accuracies with which FFRs are classified should reflect this modulation *via* improving (enhancement) or declining (suppression) classification accuracy.

A MATLAB-constructed linear support vector machine (SVM; Cristianini and Shawe-Taylor, 2000) was used to classify pre- and post-training FFR_{SWS} for test and control groups following the general procedures described by Xie et al. (2019). We first epoched all subtracted FFR_{SWS} waveforms from 0 to 380 ms and used these 1,900 amplitude-by-time points as linear SVM input features. The model outputs were stimulus type (/bɔ/, /bu/, and /bo/). Because standard linear SVM can only classify data into binary classes, a one-against-one strategy was used. In this approach, the linear SVM constructs N(N-1)/2 classifiers, where N is the number of classes; $N = 3$ in this experiment, as three SWS stimuli were used. After FFR classification is performed on all possible pairwise combinations, the class with the highest accuracy is used as the classification label.

The model was cross-validated using a three-fold approach that was repeated 2,500 times (see Xie et al., 2019; Figure 1). For each iteration of the linear SVM classifier, participants were randomly and equally divided into 3 groups (or folds). A "leave-one-out" strategy then used two of the three-folds to train the classifier. After training the classifier, the held-out fold was used as test data. This was repeated within each iteration such that each fold was held-out as the test data and the other two-folds were used for training the classifier. The average classifier accuracy across cross-validations was calculated for each iteration. Outcomes of the 2,500 iterations were also used to create grand total cross-validation accuracies as well as a distribution of accuracies. A null distribution of model accuracies was also generated using the steps above, with the exception that model outputs (i.e., stimulus labels) were randomly assigned to FFR_{SWS} inputs on each iteration of the loop. Statistical significance of "true" classifier performance was determined using $p = (a + 1)/(n + 1)$, where a denotes the number of observations from the null classification distribution that surpasses the median of the "true" distribution and n is the total number of observations comprising the null distribution (Phipson and Smyth, 2010, as cited in Xie et al., 2019). The same equation was also used to test whether pre- and post-training FFR_{SWS} classification accuracy distributions were significantly different for test and control groups.

RESULTS

Test Group Training- and Testing-Phase Behavioral Performance

Modeled accuracy and reaction times for training and test phases are plotted in **Figure 2**. For the training phase, the mixed effects logistic regression model containing training time and stimulus type as predictors was statistically significant [$X^2(2) = 37.31$, $p < 0.001$]. When holding stimulus type constant, the odds of a correct response increased by 3% [95% CI (0.13, 0.47)] for a one-unit increase in trials. When holding trial count constant, the accuracy decreased by 6% [95% CI (−0.76, −0.41)] when changing from /bɔ/ to /bu/ and /bu/ to /bo/. The multiple linear regression model evaluating reaction time suggested that training time and stimuli explained 9% of the variance, [$R^2 = 0.09$, $F_{(2,672)} = 33{,}28$, $p < 0.001$]. When holding stimulus type constant, training time significantly predicted reaction time, $\beta = -7.73$, $t = -8.09$, $p < 0.001$, suggesting the reaction time decreased when the training time increased. When holding training time constant, stimulus type did not significantly predict reaction time ($\beta = 15.62$, $t = 0.62$, $p = 0.54$).

For the test phase, the mixed effects logistic regression model containing test time and stimuli as predictors was not statistically significant [$X^2(2) = 5.38$, $p = 0.07$]. The results of multiple linear regression revealed that testing time and stimuli explained

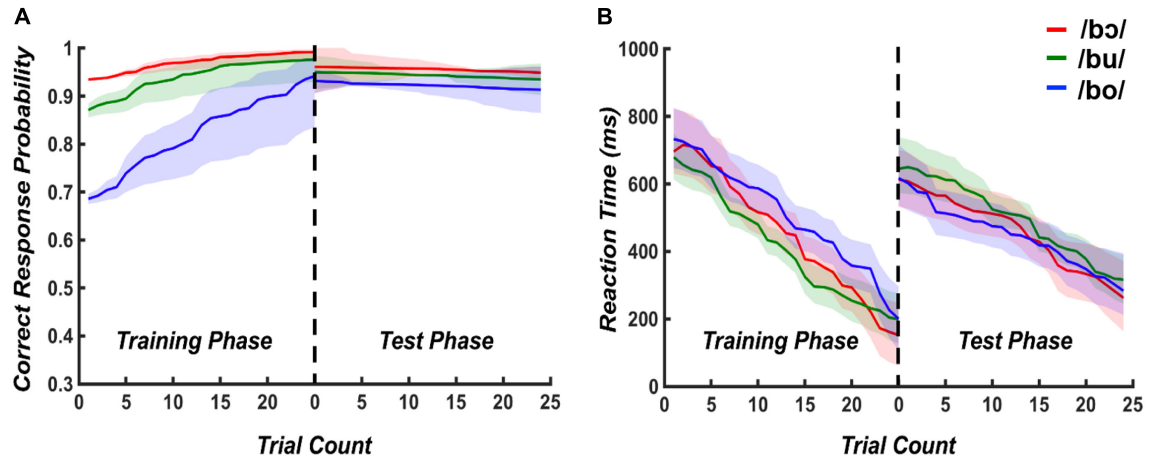

FIGURE 2 | Modeled accuracy and reaction times for training and testing phases. Logistic and multiple linear regression model outputs using stimulus type and trial count as predictors were used to plot predicted accuracy **(A)** and reaction times **(B)**, respectively. Shading indicates 95% confidence intervals of each modeled response.

3% of the variance, [R^2 = 0.03, $F_{(2,672)}$ = 12,8, $p < 0.001$]. When stimulus type was held constant, testing time significantly predicted reaction time, $\beta = -4.98$, $t = -5.06$, $p < 0.001$. However, there is no significant prediction of stimuli on reaction time ($\beta = 12.66$, $t = 0.95$, $p = 0.34$). These results collectively suggest that accuracy improved with more exposure to the stimuli in the training phase, with /bɔ/ and /bu/ being more rapidly attained than /bo/. Further, all stimuli were discriminated with a high level of accuracy during the test phase. Irrespective of the target stimulus, reaction time similarly decreased during training and testing phases.

Stimulus-to-Response Cross-Correlations

Initial stimulus-to-response calculations for pre- and post-training FFR_{SWS} showed sharply peaked cross-correlation functions between /bu/ and /bo/ and their respective SWS stimulus waveforms; because the stimulus waveforms are dominated by the F1 component, these results indicated strong neural phase locking to F1. In contrast, cross-correlations for the /bɔ/ were poor despite these FFR_{SWS} waveforms being highly periodic. Spectrographic analysis of the average FFR_{SWS} to /bɔ/ demonstrated that the neural response was in fact phase-locked to the quadratic distortion product (F2–F1) instead of F1. The F2–F1 distortion product is mechanically initiated by interactions between F2 and F1 traveling waves on the basilar membrane, and the nervous system can phase lock to this and other distortions as it would to acoustically-delivered stimuli of the same frequency (e.g., Siegel et al., 1982; Smith et al., 2017). Because F2–F1 is not present in the acoustic stimulus, the neural response does not bear a resemblance to the stimulus. To determine whether F2–F1 frequency tracking for /bɔ/ SWS_{FFR} were of neural origin, we approximated an F2–F1 "stimulus" waveform by taking the analytic envelope of the original SWS stimulus and band-passing it between 100 and 2,400 Hz. Cross-correlations were then rerun between the F2–F1 stimulus waveform and /bɔ/ SWS_{FFR}. With this adjustment, /bɔ/ SWS_{FFR} cross-correlation functions demonstrated sharp peaks similar to the other responses.

Results of the cross-correlation analyses for test and control groups are shown in **Figure 3**. Multivariate analysis showed a significant effect of stimulus type on both latency and cross-correlation strength across groups and training status, [Wilks' Lamda = 0.33, $F_{(4,62)}$ = 11.42, $p < 0.001$, $\eta^2 = 0.42$], suggesting that stimulus type affected cross-correlation strength and latencies of neural responses. The effect size, calculated using eta squared, indicated that this stimulus type effect accounted for 42% of the variance in cross-correlation strength and latency. Moreover, there was a significant effect of training status between test and control groups across stimuli, [Wilks' Lamda = 0.66, $F_{(2,15)}$ = 3.82, $p < 0.05$, $\eta^2 = 0.34$], suggesting that the interaction of training status and group affected the cross-correlation strength and latency of neural responses. The effect size, calculated using eta squared, indicated the interaction of training status and groups effect accounted for 34% of the variance in cross-correlation strength and latency. However, there is no significant stimuli*group [$F_{(4,\ 62)} = 0.12$, $p = 0.98$, $\eta^2 = 0.01$], training status [$F_{(2,\ 15)} = 1.94$, $p = 0.18$, $\eta^2 = 0.21$], stimuli*training [$F_{(4,\ 62)} = 0.59$, $p = 0.67$, $\eta^2 = 0.03$], stimuli*training*group [$F_{(4,\ 62)} = 0.12$, $p = 0.98$, $\eta^2 = 0.01$] effect on cross-correlation strength and latency [$F_{(2,\ 7)}$, $p = 0.49$, $\eta^2 = 0.19$].

Univariate tests were used to further examine the effects on latency and cross-correlation strength. These results show a significant stimulus effect on latency [Greenhouse-Geisser = 57.72, $F_{(1.88,30.08)}$ = 15.93, $p < 0.001$, $\eta^2 = 0.50$] and cross-correlation strength [Greenhouse-Geisser = 0.62, $F_{(1.99,31.87)}$ = 11, $p < 0.001$, $\eta^2 = 0.41$]. Moreover, there was a significant effect of the interaction between training status and groups on cross-correlation strength [Greenhouse-Geisser = 0.02, $F_{(1,1.68)}$ = 4.62, $p < 0.05$, $\eta^2 = 0.22$]. Within-subjects contrasts showed that latency in the /bu/ condition was significantly

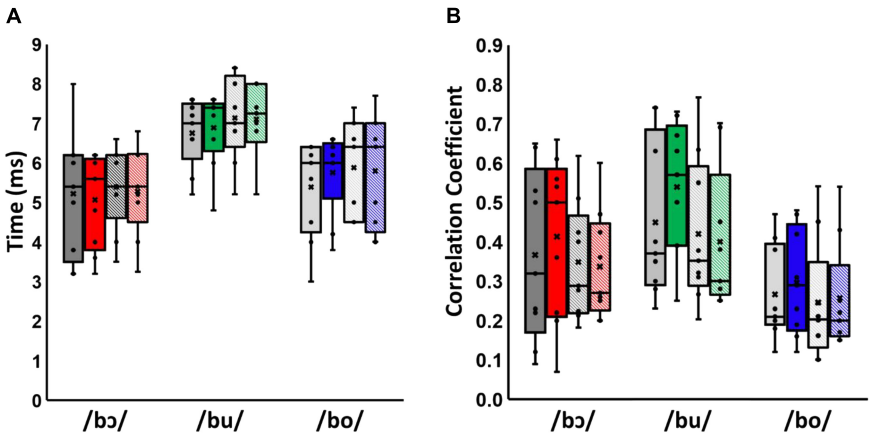

FIGURE 3 | Latency **(A)** and cross-correlation strength **(B)** for /bɔ/, /bu/, and /bo/ FFR_{SWS}. Pre-training responses are shown in gray and post-training responses are shown in color. Test group responses are solid-filled bars, whereas control group responses are cross-hatched. Means and medians are denoted by Xs and horizontal lines, respectively. Note that latencies and cross-correlations for /bɔ/ were calculated using the F2–F1 waveform as the "stimulus," as described in the text.

higher than /bɔ/ [$F_{(1,\ 16)} = 39.64$, $p < 0.001$, $\eta^2 = 0.71$] and /bo/ conditions [$F_{(1,\ 16)} = 14.59$, $p < 0.01$, $\eta^2 = 0.48$]. Additionally, cross-correlation strength in /bu/ was higher than /bo/ [$F_{(1,\ 16)} = 23.29$, $p < 0.001$, $\eta^2 = 0.59$]. Lastly, there was a significantly higher cross-correlation after training than before training in test group [$F_{(1,\ 16)} = 4.61$, $p < 0.05$, $\eta^2 = 0.22$]; this significant difference is larger than the pre- and post-training difference in control group.

Pre- and Post-training FFR_{SWS} Fourier Analyzers

FFR_{SWS} amplitudes at F1 and F2 were calculated over the duration of the response using FAs. These calculations produced a single amplitude estimate representing the strength of neural phase locking to the stimulus feature of interest over the entire duration of the stimulus. FFR_{SWS} to /bu/ and /bo/ produced measurable F1 responses above the noise floor for all subjects in pre- and post-training measurements. The issue described above regarding neural phase locking to F2–F1 in /bɔ/ FFR_{SWS} also impacted our initial FA calculations for the /bɔ/ stimulus such that F1 was not robustly represented. Consequently, we used the F2–F1 "stimulus" waveform to create an F2–F1 frequency track for /bɔ/ responses, using identical procedures described in the method section. Using this approach yielded measurable F2–F1 neural responses in every participant for /bɔ/ in pre- and post-training measurements. The following analyses focus on F1 amplitudes for /bu/ and /bo/ and F2–F1 amplitudes for /bɔ/. Because F2 was only measurable in <25% of responses, we did not further analyze these components.

Figure 4 depicts mean FFR_{SWS} waveforms and spectrograms for pre- and post-training test group responses as well as test and control group FA results for each stimulus. All post-training FFR_{SWS} responses are larger in amplitude than pre-training responses for the test group, which can be seen in waveform (**Figure 4A**) and spectrographic (**Figure 4B**) representations. Examination of the FA results for the test group (**Figure 4C**) demonstrates that these enhancements are at the frequency of interest only and are not observed in the adjacent noise bins. By comparison, FFR_{SWS} enhancements are not apparent in the control group FA responses.

The effect of training status and group on FA amplitudes was analyzed using a two-way MANOVA with repeated measures. This analysis showed an interaction of training status and group [$F_{(5,\ 12)} = 3.09$, $p = 0.05$, $\eta^2 = 0.56$] on FA amplitude. Univariate tests revealed significant interactions between training status and group for /bɔ/ [Greenhouse-Geisser = 0.002, $F_{(1,16)} = 6.72$, $p < 0.05$, $\eta^2 = 0.30$], /bu/ [Greenhouse-Geisser = 0.004, $F_{(1,16)} = 8.44$, $p < 0.05$, $\eta^2 = 0.34$], and /bo/ [Greenhouse-Geisser = 0.002, $F_{(1,16)} = 9.96$, $p < 0.01$, $\eta^2 = 0.38$] FA amplitudes. Pre- and post-training differences between test and control groups revealed that /bɔ/ [$F_{(1,\ 16)} = 6.73$, $p < 0.05$, $\eta^2 = 0.30$], /bu/ [$F_{(1,\ 16)} = 8.44$, $p < 0.05$, $\eta^2 = 0.35$], and /bo/ [$F_{(1,\ 16)} = 9.96$, $p < 0.01$, $\eta^2 = 0.38$] FA amplitudes were significantly higher after training in test group but not in the control group.

Machine Learning Classification of FFR_{SWS}

Average linear SVM classification accuracies for test and control groups are depicted for pre-training and post-training FFR_{SWS} in the confusion matrices of **Figure 5**. For the test group, pre-training classification accuracy was poorer for each stimulus classifier relative to post-training accuracy, whereas control group pre- and post-training classification results do not follow a clear pattern. Overall classification accuracy distributions representing all 2,500 iterations are depicted in the 3D histogram plots, as are empirical null distributions generated by randomly shuffling classifier outputs (i.e., response labels) for each iteration. Pre-training ($p < 0.001$) and post-training ($p < 0.001$) FFR_{SWS} classification was significantly above the null distribution for test and control groups, as determined using $p = (a + 1)/(n + 1)$. In the test group, post-training classification was significantly higher ($p < 0.01$) than pre-training classification using the

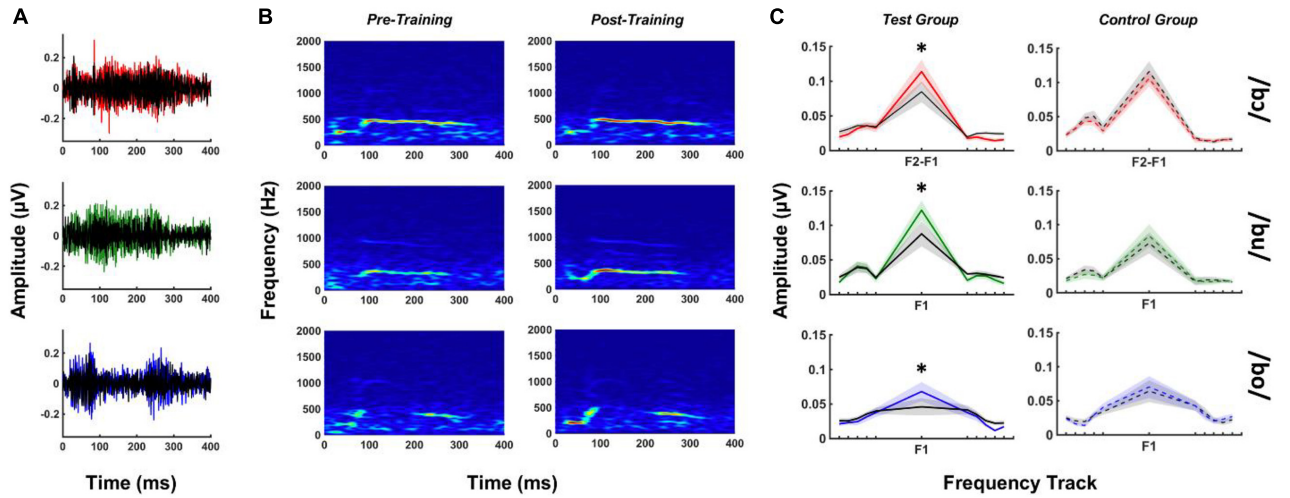

FIGURE 4 | FFR_{SWS} pre- and post-training waveforms **(A)**, spectrograms **(B)**, and FA results **(C)** for each stimulus. **(A,B)** Are from test group data only, whereas **(C)** includes test and control FA results for comparison. Colored (red, green, and blue) waveforms and FA spectra represent post-training responses, whereas black waveforms are their pre-training counterparts. FA center frequencies (F2–F1 or F1) are denoted for each FA plot; noise bins starting at ± 5 Hz relative to the frequency of interest are indicated by peripheral tick marks on the *x*-axis (shading = SEM). All results represented subtracted waveforms. $^*p < 0.05$.

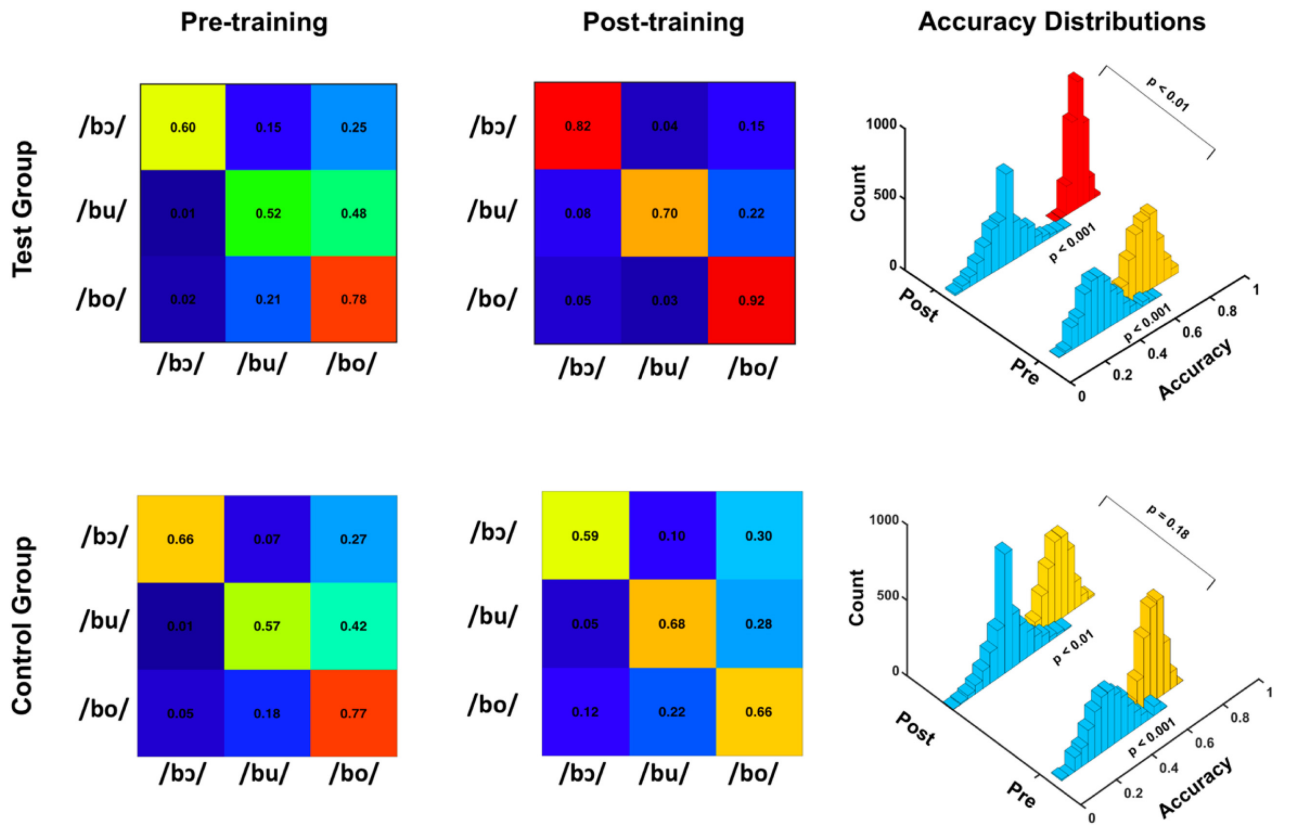

FIGURE 5 | Pre-training and post-training classification accuracies for test (top) and control (bottom) groups. Confusion matrices on the left demonstrate linear SVM classification accuracies for pre-training FFR_{SWS}, whereas the confusion matrices in the center demonstrate post-training accuracies. 3D histograms on the right show empirical null distributions (teal bins) and pre- and post-training average classification accuracy distributions for test and control groups. Pre- and post-training distributions were significantly above the null distribution for test and control groups. Additionally, the test group post-training distribution was significantly more accurate than the pre-training distribution.

same equation. In contrast, pre- and post-training classification were not different in the control group ($p = 0.18$). Collectively, these results indicate that classification of pre- and post-training FFR_{SWS} was significantly above chance for test and control groups; however, post-training data were classified with significantly higher accuracy than pre-training data in the test group compared to the control group.

DISCUSSION

To our knowledge, this is the first study using FFR_{SWS} to examine training or context effects in the auditory brainstem. Utilizing SWS in this context allowed for direct comparisons between pre- and post-training FFR_{SWS} evoked by identical, acoustically-sparse speech stimuli that initiated neural responses from more caudal subcortical sources than stimuli used in previous reports. Because most listeners do not hear SWS as speech unless they are provided with additional instruction (Remez et al., 1981; Möttönen et al., 2006), pre-training neural representation of SWS theoretically offers a glimpse into bottom-up auditory processing of the "naïve" auditory nervous system. When additional instruction or context is provided to listeners regarding SWS, they often attain a high level of comprehension in a brief period of time or, in some cases, immediately (Remez et al., 1981; Möttönen et al., 2006). Thus, post-training FFR_{SWS} may offer insight into how rapidly and potently the auditory brainstem can be functionally modulated *via* the efferent system when speech comprehension networks are engaged.

Our behavioral results suggest that SWS stimuli were rapidly attained in the training trials, albeit at slightly different rates. For example, /bɔ/ and /bu/ discrimination reached peak performance within relatively few trials, whereas /bo/ required more training before responses were consistently accurate. This pattern suggests that /bo/ was initially more difficult to discriminate than /bɔ/ and /bu/, which may simply be explained by acoustical differences (i.e., /bo/ and /bu/ F1 and F2 contours are more similar than /bɔ/ and were spaced such that they were unlikely to generate strong distortion products; see **Figure 1**). An additional revelation from our FFR_{SWS} data was that participants may have benefitted from hearing the F2–F1 distortion product created by the /bɔ/ stimulus. The F2–F1 frequency is generated by mechanical interaction on the basilar membrane and "feeds forward" into the auditory nervous system, as do other distortion products (Siegel et al., 1982; Chertoff et al., 1992; Dhar et al., 2009; Smith et al., 2017), effectively converting a dynamic two-tone stimulus into a perceptually richer input (Goldstein et al., 1978). Because we did not assess psychophysical weighting of the F2–F1 cue, it is not clear whether its post-training enhancement in the FFR was a consequence of direct attention to this cue or a gross upscaling of any auditory stimuli relevant to the perceptual task.

Despite stimulus-related differences in behavioral training results, we observed that all FFR_{SWS} in the test group were enhanced in the post-training phase relative to the pre-training phase and compared to a control group. This was indicated in larger FA amplitudes of F1 (for /bo/ and /bu/) and F2–F1 (for /bɔ/), as well as higher machine learning classification accuracy in the post-training phase. Importantly, these differences were due to FFR_{SWS} amplitude enhancement and not differences in residual noise between pre- and post-training responses, as evidenced by the FA noise tracks (**Figure 4**). There are multiple potential explanations for the observed FFR_{SWS} enhancements in the test group. First, the rapid perceptual shift from non-speech to speech may have engaged speech comprehension networks originating in frontal cortex and extending through auditory cortex and brain stem (Davis and Johnsrude, 2003; Eisner et al., 2010; Hervais-Adelman et al., 2012; Bidelman et al., 2018a, 2019; Khoshkhoo et al., 2018). That FFR_{SWS} enhancements reflect a more immediate *online* context shift and not short-term training *per se* is supported by a few congruent observations in our behavioral and neurophysiologic data. For example, perceptual shifts appear to have occurred quickly for /bɔ/ and /bu/ stimuli, whereas /bo/ required more exposure trials before it was attained. These results suggest that perceptual salience of the context shift may have differed slightly across stimuli. The size of FA enhancements and within-class changes in linear SVM classification accuracy between pre- and post-training trials mirror the behavioral results: more immediate behavioral SWS identification was associated with larger FA enhancements and greater changes in classification accuracy post-training. It is also notable that the enhancements observed in the present study appeared earlier than many reports on FFR training effects, which required multiple hours to days of training (e.g., Song et al., 2008; Carcagno and Plack, 2011). This may be related to the fact that SWS was initially processed as a completely different class of stimulus (e.g., uncorrelated "whistles") before being recognized as speech, whereas participants in previous studies were aware from experiment onset that they were hearing speech or music stimuli. Further, SWS forces listeners to focus on a minimal number of cues (F1, F2, and/or F2–F1), whereas speech and music pitch may be determined in a variety of ways, such as listening to resolved and/or unresolved harmonics (e.g., Laudanski et al., 2014); therefore, attention may be allocated to different channels of information summating to produce the FFR. A limitation of our approach, which does not allow us to resolve single-trial FFR_{SWS}, is that we cannot delineate whether the observed enhancements are related to online or short-term changes following the perceptual shift.

A control group was used in the present study to examine whether post-training vs. pre-training differences were simply a result of more exposure to the stimuli during the experimental protocol. Our results suggest that this is not the case, as the control group responses were not enhanced "post-training" relative to "pre-training." These results comport with the multiple studies that have demonstrated high test-retest reliability of FFR amplitudes within and between *passive* test sessions (e.g., Song et al., 2011; Bidelman et al., 2018b; Easwar et al., 2020).

Future studies will examine afferent-efferent connectivity using similar SWS stimuli to examine the time course and neural substrates involved in perceptual shifts and/or training effects reported here. Because of the simple, sinusoidal nature of SWS, it may also be possible to measure simultaneous stimulus frequency otoacoustic emissions in addition to neural responses from the

brainstem and cortex. Such an approach would allow for context or training effects to be studied from cochlea to cortex using the same stimuli.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the University of Texas at Austin IRB. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work, and approved it for publication.

FUNDING

This research was funded by the National Institutes of Health, National Institute on Deafness and other Communication Disorders (K01DC017192).

REFERENCES

Aiken, S. J., and Picton, T. W. (2006). Envelope following responses to natural vowels. *Audiol. Neurotol.* 11, 213–232. doi: 10.1159/000092589

Aiken, S. J., and Picton, T. W. (2008). Envelope and spectral frequency-following responses to vowel sounds. *Hear. Res.* 245, 35–47. doi: 10.1016/j.heares.2008.08.004

Atiani, S., David, S. V., Elgueda, D., Locastro, M., Radtke-Schuller, S., Shamma, S. A., et al. (2014). Emergent selectivity for task-relevant stimuli in higher-order auditory cortex. *Neuron* 82, 486–499. doi: 10.1016/j.neuron.2014.02.029

Bajo, V. M., Nodal, F. R., Moore, D. R., and King, A. J. (2010). The descending corticocollicular pathway mediates learning-induced auditory plasticity. *Nat. Neurosci.* 13, 253–260. doi: 10.1038/nn.2466

Barker, J., and Cooke, M. (1999). Is the sine-wave speech cocktail party worth attending? *Speech Commun.* 27, 159–174. doi: 10.1016/s0167-6393(98)00081-8

Bidelman, G., and Powers, L. (2018). Response properties of the human frequency-following response (FFR) to speech and non-speech sounds: level dependence, adaptation and phase-locking limits. *Int. J. Audiol.* 57, 665–672. doi: 10.1080/14992027.2018.1470338

Bidelman, G. M. (2015). Multichannel recordings of the human brainstem frequency-following response: scalp topography, source generators, and distinctions from the transient ABR. *Hear. Res.* 323, 68–80. doi: 10.1016/j.heares.2015.01.011

Bidelman, G. M. (2018). Subcortical sources dominate the neuroelectric auditory frequency-following response to speech. *Neuroimage* 175, 56–69. doi: 10.1016/j.neuroimage.2018.03.060

Bidelman, G. M., and Krishnan, A. (2009). Neural correlates of consonance, dissonance, and the hierarchy of musical pitch in the human brainstem. *J. Neurosci.* 29, 13165–13171. doi: 10.1523/JNEUROSCI.3900-09.2009

Bidelman, G. M., Nelms, C., and Bhagat, S. P. (2016). Musical experience sharpens human cochlear tuning. *Hear. Res.* 335, 40–46. doi: 10.1016/j.heares.2016.02.012

Bidelman, G. M., Davis, M. K., and Pridgen, M. H. (2018a). Brainstem-cortical functional connectivity for speech is differentially challenged by noise and reverberation. *Hear. Res.* 367, 149–160. doi: 10.1016/j.heares.2018.05.018

Bidelman, G. M., Pousson, M., Dugas, C., and Fehrenbach, A. (2018b). Test–retest reliability of dual-recorded brainstem versus cortical auditory-evoked potentials to speech. *J. Am. Acad. Audiol.* 29, 164–174. doi: 10.3766/jaaa.16167

Bidelman, G. M., Price, C. N., Shen, D., Arnott, S. R., and Alain, C. (2019). Afferent-efferent connectivity between auditory brainstem and cortex accounts for poorer speech-in-noise comprehension in older adults. *Hear. Res.* 382:107795. doi: 10.1016/j.heares.2019.107795

Bidelman, G. M., Schneider, A. D., Heitzmann, V. R., and Bhagat, S. P. (2017). Musicianship enhances ipsilateral and contralateral efferent gain control to the cochlea. *Hear. Res.* 344, 275–283. doi: 10.1016/j.heares.2016.12.001

Bidelman, G. M., Schug, J. M., Jennings, S. G., and Bhagat, S. P. (2014). Psychophysical auditory filter estimates reveal sharper cochlear tuning in musicians. *J. Acoust. Soc. Am.* 136, EL33–EL39.

Boersma, P. (2009). *Praat: Doing Phonetics by Computer (Version 5.1. 05)*. Available online at: http://www.praat.org/ (accessed August 20, 2018).

Carcagno, S., and Plack, C. J. (2011). Subcortical plasticity following perceptual learning in a pitch discrimination task. *J. Assoc. Res. Otolaryngol.* 12, 89–100. doi: 10.1007/s10162-010-0236-1

Chandrasekaran, B., and Kraus, N. (2010). The scalp-recorded brainstem response to speech: neural origins and plasticity. *Psychophysiology* 47, 236–246. doi: 10.1111/j.1469-8986.2009.00928.x

Chandrasekaran, B., Kraus, N., and Wong, P. C. (2012). Human inferior colliculus activity relates to individual differences in spoken language learning. *J. Neurophysiol.* 107, 1325–1336. doi: 10.1152/jn.00923.2011

Chertoff, M. E., Hecox, K. E., and Goldstein, R. (1992). Auditory distortion products measured with averaged auditory evoked potentials. *J. Speech Lang. Hear. Res.* 35, 157–166. doi: 10.1044/jshr.3501.157

Choi, J. M., Purcell, D. W., Coyne, J. A. M., and Aiken, S. J. (2013). Envelope following responses elicited by English sentences. *Ear Hear.* 34, 637–650. doi: 10.1097/AUD.0b013e31828e4dad

Coffey, E. B., Herholz, S. C., Chepesiuk, A. M., Baillet, S., and Zatorre, R. J. (2016). Cortical contributions to the auditory frequency-following response revealed by MEG. *Nat. Commun.* 7:11070. doi: 10.1038/ncomms11070

Coffey, E. B., Nicol, T., White-Schwoch, T., Chandrasekaran, B., Krizman, J., Skoe, E., et al. (2019). Evolving perspectives on the sources of the frequency-following response. *Nat. Commun.* 10:5036. doi: 10.1038/s41467-019-13003-w

Cohen, J., Cohen, P., West, S. G., and Aiken, L. S. (2013). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*. London: Routledge.

Connolly, J. F., Aubry, K., McGillivary, N., and Scott, D. W. (1989). Human brainstem auditory evoked potentials fail to provide evidence of efferent modulation of auditory input during attentional tasks. *Psychophysiology* 26, 292–303. doi: 10.1111/j.1469-8986.1989.tb01920.x

Cristianini, N., and Shawe-Taylor, J. (2000). *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*. Cambridge: Cambridge university press.

Darwin, C. (2003). *Sine-Wave Speech Produced Automatically Using a Script for the PRAAT Program*. Available online at: http://www.lifesci.sussex.ac.uk/home/Chris_Darwin/SWS/ (accessed December 8, 2011).

Davis, M. H., and Johnsrude, I. S. (2003). Hierarchical processing in spoken language comprehension. *J. Neurosci.* 23, 3423–3431. doi: 10.1523/JNEUROSCI.23-08-03423.2003

de Boer, J., and Thornton, A. R. D. (2008). Neural correlates of perceptual learning in the auditory brainstem: efferent activity predicts and reflects improvement at a speech-in-noise discrimination task. *J. Neurosci.* 28, 4929–4937. doi: 10.1523/JNEUROSCI.0902-08.2008

Delano, P. H., and Elgoyhen, A. B. (2016). Editorial: auditory efferent system: new insights from cortex to cochlea. *Front. Syst. Neurosci.* 10:50. doi: 10.3389/fnsys.2016.00050

Dhar, S., Abel, R., Hornickel, J., Nicol, T., Skoe, E., Zhao, W., et al. (2009). Exploring the relationship between physiological measures of cochlear and brainstem function. *Clin. Neurophysiol.* 120, 959–966. doi: 10.1016/j.clinph.2009.02.172

Dragicevic, C. D., Aedo, C., León, A., Bowen, M., Jara, N., Terreros, G., et al. (2015). The olivocochlear reflex strength and cochlear sensitivity are independently modulated by auditory cortex microstimulation. *J. Assoc. Res. Otolaryngol.* 16, 223–240. doi: 10.1007/s10162-015-0509-9

Easwar, V., Scollie, S., Aiken, S., and Purcell, D. (2020). Test-Retest variability in the characteristics of envelope following responses evoked by speech stimuli. *Ear Hear.* 41, 150–164. doi: 10.1097/AUD.0000000000000739

Eggermont, J. J. (2017). Ups and downs in 75 years of electrocochleography. *Front. Syst. Neurosci.* 11:2. doi: 10.3389/fnsys.2017.00002

Eisner, F., McGettigan, C., Faulkner, A., Rosen, S., and Scott, S. K. (2010). Inferior frontal gyrus activation predicts individual differences in perceptual learning of cochlear-implant simulations. *J. Neurosci.* 30, 7179–7186. doi: 10.1523/JNEUROSCI.4040-09.2010

Etard, O., Kegler, M., Braiman, C., Forte, A. E., and Reichenbach, T. (2019). Decoding of selective attention to continuous speech from the human auditory brainstem response. *Neuroimage* 200, 1–11. doi: 10.1016/j.neuroimage.2019.06.029

Forte, A. E., Etard, O., and Reichenbach, T. (2017). The human auditory brainstem response to running speech reveals a subcortical mechanism for selective attention. elife 6:e27203. doi: 10.7554/eLife.27203

Francis, N. A., Zhao, W., and Guinan, J. J. Jr. (2018). Auditory attention reduced ear-canal noise in humans by reducing subject motion, not by medial olivocochlear efferent inhibition: implications for measuring otoacoustic emissions during a behavioral task. *Front. Syst. Neurosci.* 12:42. doi: 10.3389/fnsys.2018.00042I

Galbraith, G. C., Arbagey, P. W., Branski, R., Comerci, N., and Rector, P. M. (1995). Intelligible speech encoded in the human brain stem frequency-following response. *Neuroreport* 6, 2363–2367. doi: 10.1097/00001756-199511270-00021

Galbraith, G. C., and Arroyo, C. (1993). Selective attention and brainstem frequency-following responses. *Biol. Psychol.* 37, 3–22. doi: 10.1016/0301-0511(93)90024-3

Galbraith, G. C., Bagasan, B., and Sulahian, J. (2001). Brainstem frequency-following response recorded from one vertical and three horizontal electrode derivations. *Percept. Motor Skills* 92, 99–106. doi: 10.2466/pms.2001.92.1.99

Galbraith, G. C., Bhuta, S. M., Choate, A. K., Kitahara, J. M., and Mullen, T. A. Jr. (1998). Brain stem frequency-following response to dichotic vowels during attention. *Neuroreport* 9, 1889–1893. doi: 10.1097/00001756-199806010-00041

Galbraith, G. C., Olfman, D. M., and Huffman, T. M. (2003). Selective attention affects human brain stem frequency-following response. *Neuroreport* 14, 735–738. doi: 10.1097/00001756-200304150-00015

Gao, E., and Suga, N. (1998). Plasticity of midbrain auditory frequency map mediated by the corticofugal system in bat. *Proc. Natl. Acad. Sci. U.S.A.* 95, 12663–12670. doi: 10.1038/255

Gao, E., and Suga, N. (2000). Experience-dependent plasticity in the auditory cortex and the inferior colliculus of bats: role of the corticofugal system. *Proc. Natl. Acad. Sci. U.S.A.* 97, 8081–8086. doi: 10.1073/pnas.97.14.8081

Gardi, J., Merzenich, M., and McKean, C. (1979). Origins of the scalp-recorded frequency-following response in the cat. *Audiology* 18, 353–380. doi: 10.3109/00206097909070062

Goldstein, J. L., Buchsbaum, G., and Furst, M. (1978). Compatibility between psychophysical and physiological measurements of aural combination tones. *J. Acoust. Soc. Am.* 63, 474–485. doi: 10.1121/1.381739

Gregory, S. D., Heath, J. A., and Rosenberg, M. E. (1989). Does selective attention influence the brain-stem auditory evoked potential? *Electroencephalogr. Clin. Neurophysiol.* 73, 557–560. doi: 10.1016/0013-4694(89)90266-6

Hackley, S. A., Woldorff, M., and Hillyard, S. A. (1990). Cross-modal selective attention effects on retinal, myogenic, brainstem, and cerebral evoked potentials. *Psychophysiology* 27, 195–208. doi: 10.1111/j.1469-8986.1990.tb00370.x

Hairston, W. D., Letowski, T. R., and McDowell, K. (2013). Task-related suppression of the brainstem frequency following response. *PLoS One* 8:e55215. doi: 10.1371/journal.pone.0055215

Hernandez-Peon, R., Scherrer, H., and Jouvet, M. (1956). Modification of electric activity in cochlear nucleus during attention in unanesthetized cats. *Science* 123, 331–332. doi: 10.1126/science.123.3191.331

Hernandez-Perez, H., Mikiel-Hunter, J., McAlpine, D., Dhar, S., Boothalingam, S., Monaghan, J. J., et al. (2021). Perceptual gating of a brainstem reflex facilitates speech understanding in human listeners. *bioRxiv* [Preprint]. doi: 10.1101/2020.05.31.115444

Hervais-Adelman, A. G., Carlyon, R. P., Johnsrude, I. S., and Davis, M. H. (2012). Brain regions recruited for the effortful comprehension of noise-vocoded words. *Lang. Cogn. Proces.* 27, 1145–1166. doi: 10.1080/01690965.2012.662280

Holdgraf, C. R., Rieger, J. W., Micheli, C., Martin, S., Knight, R. T., and Theunissen, F. E. (2017). Encoding and decoding models in cognitive electrophysiology. *Front. Syst. Neurosci.* 11:61. doi: 10.3389/fnsys.2017.00061

Jedrzejczak, W. W., Milner, R., Ganc, M., Pilka, E., and Skarzynski, H. (2020). No change in medial olivocochlear efferent activity during an auditory or visual task: dual evidence from otoacoustic emissions and event-related potentials. *Brain Sci.* 10:894. doi: 10.3390/brainsci10110894

Ji, W., Gao, E., and Suga, N. (2001). Effects of acetylcholine and atropine on plasticity of central auditory neurons caused by conditioning in bats. *J. Neurophysiol.* 86, 211–225. doi: 10.1152/jn.2001.86.1.211

Khoshkhoo, S., Leonard, M. K., Mesgarani, N., and Chang, E. F. (2018). Neural correlates of sine-wave speech intelligibility in human frontal and temporal cortex. *Brain Lang.* 187, 83–91. doi: 10.1016/j.bandl.2018.01.007

Kraus, N., and Chandrasekaran, B. (2010). Music training for the development of auditory skills. *Nat. Rev. Neurosci.* 11, 599–605. doi: 10.1038/nrn2882

Kraus, N., and White-Schwoch, T. (2015). Unraveling the biology of auditory learning: a cognitive–sensorimotor–reward framework. *Trends Cogn. Sci.* 19, 642–654. doi: 10.1016/j.tics.2015.08.017

Krishnan, A., and Gandour, J. T. (2009). The role of the auditory brainstem in processing linguistically-relevant pitch patterns. *Brain Lang.* 110, 135–148. doi: 10.1016/j.bandl.2009.03.005

Krishnan, A., Xu, Y., Gandour, J., and Cariani, P. (2005). Encoding of pitch in the human brainstem is sensitive to language experience. *Cogn. Brain Res.* 25, 161–168. doi: 10.1016/j.cogbrainres.2005.05.004

Laudanski, J., Zheng, Y., and Brette, R. (2014). A structural theory of pitch. *eNeuro* 1:ENEURO.33-ENEURO.14. doi: 10.1523/ENEURO.0033-14.2014

Lauer, A. M., Jimenez, S. V., and Delano, P. H. (2021). Olivocochlear efferent effects on perception and behavior. *Hear. Res.* 108207. doi: 10.1016/j.heares.2021.108207

Lehmann, A., and Schönwiesner, M. (2014). Selective attention modulates human auditory brainstem responses: relative contributions of frequency and spatial cues. *PLoS One* 9:e85442. doi: 10.1371/journal.pone.0085442

León, A., Elgueda, D., Silva, M. A., Hamamé, C. M., and Delano, P. H. (2012). Auditory cortex basal activity modulates cochlear responses in chinchillas. *PLoS One* 7:e36203. doi: 10.1371/journal.pone.0036203

Lichtenhan, J. T., Cooper, N. P., and Guinan, J. J. (2013). A new auditory threshold estimation technique for low frequencies: proof of concept. *Ear Hear.* 34:42. doi: 10.1097/AUD.0b013e31825f9bd3

Lindblom, B. (1990). "Explaining phonetic variation: a sketch of the H&H theory," in *Speech Production and Speech Modelling*, eds W. J. Hardcastle and A. Marchal (Dordrecht: Springer), 403–439. doi: 10.1121/1.405815

Llanos, F., Xie, Z., and Chandrasekaran, B. (2017). Hidden Markov modeling of frequency-following responses to Mandarin lexical tones. *J. Neurosci. Methods* 291, 101–112. doi: 10.1016/j.jneumeth.2017.08.010

Ma, X., and Suga, N. (2001). Plasticity of bat's central auditory system evoked by focal electric stimulation of auditory and/or somatosensory cortices. *J. Neurophysiol.* 85, 1078–1087. doi: 10.1152/jn.2001.85.3.1078

Malmierca, M. S., Cristaudo, S., Pérez-González, D., and Covey, E. (2009). Stimulus-specific adaptation in the inferior colliculus of the anesthetized rat. *J. Neurosci.* 29, 5483–5493. doi: 10.1523/JNEUROSCI.4153-08.2009

May, B. J., Budelis, J., and Niparko, J. K. (2004). Behavioral studies of the olivocochlear efferent system: learning to listen in noise. *Arch. Otolaryngol. Head Neck Surg.* 130, 660–664. doi: 10.1001/archotol.130.5.660

Möttönen, R., Calvert, G. A., Jääskeläinen, I. P., Matthews, P. M., Thesen, T., Tuomainen, J., et al. (2006). Perceiving identical sounds as speech or non-speech modulates activity in the left posterior superior temporal sulcus. *Neuroimage* 30, 563–569. doi: 10.1016/j.neuroimage.2005.10.002

Perrot, X., Ryvlin, P., Isnard, J., GuÈnot, M., Catenoix, H., Fischer, C., et al. (2006). Evidence for corticofugal modulation of peripheral auditory activity in humans. *Cereb. Cortex* 16, 941–948. doi: 10.1093/cercor/bhj035

Phipson, B., and Smyth, G. K. (2010). Permutation P-values should never be zero: calculating exact P-values when permutations are randomly drawn. *Stat. Appl. Genet. Mol. Biol.* 9:39. doi: 10.2202/1544-6115.1585

Picton, T. W., and Hillyard, S. A. (1974). Human auditory evoked potentials II: effects of attention. *Electroencephalogr. Clin. Neurophysiol.* 36, 191–199. doi: 10.1016/0013-4694(74)90156-4

Picton, T. W., John, M. S., Dimitrijevic, A., and Purcell, D. (2003). Human auditory steady-state responses: respuestas auditivas de estado estable en humanos. *Int. J. Audiol.* 42, 177–219. doi: 10.3109/14992020309101316

Price, C. N., and Bidelman, G. M. (2021). Attention reinforces human corticofugal system to aid speech perception in noise. *Neuroimage* 235:118014. doi: 10.1016/j.neuroimage.2021.118014

Purcell, D. W., John, S. M., Schneider, B. A., and Picton, T. W. (2004). Human temporal auditory acuity as assessed by envelope following responses. *J. Acoust. Soc. Am.* 116, 3581–3593. doi: 10.1121/1.1798354

Remez, R., Rubin, P., Pisoni, D., and Carrell, T. (1981). Speech perception without traditional speech cues. *Science* 212, 947–949. doi: 10.1126/science.7233191

Ritter, W., Simson, R., and Vaughan, H. G. Jr. (1972). Association cortex potentials and reaction time in auditory discrimination. *Electroencephalogr. Clin. Neurophysiol.* 33, 547–555. doi: 10.1016/0013-4694(72)90245-3

Ruggles, D., Bharadwaj, H., and Shinn-Cunningham, B. G. (2012). Why middle-aged listeners have trouble hearing in everyday settings. *Curr. Biol.* 22, 1417–1422. doi: 10.1016/j.cub.2012.05.025

Russo, N. M., Nicol, T. G., Zecker, S. G., Hayes, E. A., and Kraus, N. (2005). Auditory training improves neural timing in the human brainstem. *Behav. Brain Res.* 156, 95–103. doi: 10.1016/j.bbr.2004.05.012

Sadeghian, A., Dajani, H. R., and Chan, A. D. (2015). Classification of speech-evoked brainstem responses to English vowels. *Speech Commun.* 68, 69–84. doi: 10.1016/j.specom.2015.01.003

Saiz-Alía, M., Forte, A. E., and Reichenbach, T. (2019). Individual differences in the attentional modulation of the human auditory brainstem response to speech inform on speech-in-noise deficits. *Sci. Rep.* 9:14131. doi: 10.1038/s41598-019-50773-1

Shaheen, L. A., Slee, S. J., and David, S. V. (2021). Task engagement improves neural discriminability in the auditory midbrain of the marmoset monkey. *J. Neurosci.* 41, 284–297. doi: 10.1523/JNEUROSCI.1112-20.2020

Siegel, J. H., Kim, D. O., and Molnar, C. E. (1982). Effects of altering organ of Corti on cochlear distortion products f2-f1 and 2f1-f2. *J. Neurophysiol.* 47, 303–328. doi: 10.1152/jn.1982.47.2.303

Skoe, E., Chandrasekaran, B., Spitzer, E. R., Wong, P. C., and Kraus, N. (2014). Human brainstem plasticity: the interaction of stimulus probability and auditory learning. *Neurobiol. Learn. Mem.* 109, 82–93. doi: 10.1016/j.nlm.2013.11.011

Skoe, E., and Kraus, N. (2010). Hearing it again and again: on-line subcortical plasticity in humans. *PLoS One* 5:e13645. doi: 10.1371/journal.pone.0013645

Skoe, E., Krizman, J., Spitzer, E., and Kraus, N. (2013). The auditory brainstem is a barometer of rapid auditory learning. *Neuroscience* 243, 104–114.

Slee, S. J., and David, S. V. (2015). Rapid task-related plasticity of spectrotemporal receptive fields in the auditory midbrain. *J. Neurosci.* 35, 13090–13102. doi: 10.1523/JNEUROSCI.1671-15.201

Smith, S. B., and Cone, B. (2015). The medial olivocochlear reflex in children during active listening. *Int. J. Audiol.* 54, 518–523. doi: 10.3109/14992027.2015.1008105

Smith, S. B., Ichiba, K., Velenovsky, D. S., and Cone, B. (2017). Efferent modulation of pre-neural and neural distortion products. *Hear. Res.* 356, 25–34. doi: 10.1016/j.heares.2017.10.009

Song, J. H., Nicol, T., and Kraus, N. (2011). Test–retest reliability of the speech-evoked auditory brainstem response. *Clin. Neurophysiol.* 122, 346–355.

Song, J. H., Skoe, E., Banai, K., and Kraus, N. (2012). Training to improve hearing speech in noise: biological mechanisms. *Cereb. Cortex* 22, 1180–1190. doi: 10.1093/cercor/bhr196

Song, J. H., Skoe, E., Wong, P. C., and Kraus, N. (2008). Plasticity in the adult human auditory brainstem following short-term linguistic training. *J. Cogn. Neurosci.* 20, 1892–1902. doi: 10.1162/jocn.2008.20131

Strait, D. L., and Kraus, N. (2014). Biological impact of auditory expertise across the life span: musicians as a model of auditory learning. *Hear. Res.* 308, 109–121. doi: 10.1016/j.heares.2013.08.004

Stuart, A., and Butler, A. K. (2012). Contralateral suppression of transient otoacoustic emissions and sentence recognition in noise in young adults. *J. Am. Acad. Audiol.* 23, 686–696. doi: 10.3766/jaaa.23.9.3

Suga, N., Gao, E., Zhang, Y., Ma, X., and Olsen, J. F. (2000). The corticofugal system for hearing: recent progress. *Proc. Natl. Acad. Sci. U.S.A.* 97, 11807–11814. doi: 10.1073/pnas.97.22.11807

Suga, N., Xiao, Z., Ma, X., and Ji, W. (2002). Plasticity and corticofugal modulation for hearing in adult animals. *Neuron* 36, 9–18. doi: 10.1016/s0896-6273(02)00933-9

Swaminathan, J., Krishnan, A., and Gandour, J. T. (2008). Pitch encoding in speech and nonspeech contexts in the human auditory brainstem. *Neuroreport* 19:1163. doi: 10.1097/WNR.0b013e3283088d31

Terreros, G., and Delano, P. H. (2015). Corticofugal modulation of peripheral auditory responses. *Front. Syst. Neurosci.* 9:134. doi: 10.3389/fnsys.2015.00134

Tichko, P., and Skoe, E. (2017). Frequency-dependent fine structure in the frequency-following response: the byproduct of multiple generators. *Hear. Res.* 348, 1–15. doi: 10.1016/j.heares.2017.01.014

Varghese, L., Bharadwaj, H. M., and Shinn-Cunningham, B. G. (2015). Evidence against attentional state modulating scalp-recorded auditory brainstem steady-state responses. *Brain Res.* 1626, 146–164. doi: 10.1016/j.brainres.2015.06.038

Winer, J. A. (2005). Decoding the auditory corticofugal systems. *Hear. Res.* 207, 1–9.

Wittekindt, A., Kaiser, J., and Abel, C. (2014). Attentional modulation of the inner ear: a combined otoacoustic emission and EEG study. *J. Neurosci.* 34, 9995–10002. doi: 10.1523/JNEUROSCI.4861-13.2014

Woldorff, M., Hansen, J. C., and Hillyard, S. A. (1987). Evidence for effects of selective attention in the mid-latency range of the human auditory event-related potential. *Electroencephalogr. Clin. Neurophysiol. Suppl.* 40, 146–154.

Wong, P. C., Skoe, E., Russo, N. M., Dees, T., and Kraus, N. (2007). Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nat. Neurosci.* 10, 420–422. doi: 10.1038/nn1872

Xiao, Z., and Suga, N. (2002). Modulation of cochlear hair cells by the auditory cortex in the mustached bat. *Nat. Neurosci.* 5, 57–63. doi: 10.1038/nn786

Xie, Z., Reetzke, R., and Chandrasekaran, B. (2018). Taking attention away from the auditory modality: context-dependent effects on early sensory encoding of speech. *Neuroscience* 384, 64–75. doi: 10.1016/j.neuroscience.2018.05.023

Xie, Z., Reetzke, R., and Chandrasekaran, B. (2019). Machine learning approaches to analyze speech-evoked neurophysiological responses. *J. Speech Lang. Hear. Res.* 62, 587–601. doi: 10.1044/2018_JSLHR-S-ASTM-18-0244

Yan, J., and Suga, N. (1999). Corticofugal amplification of facilitative auditory responses of subcortical combination-sensitive neurons in the mustached bat. *J. Neurophysiol.* 81, 817–824. doi: 10.1152/jn.1999.81.2.817

Yan, J., Zhang, Y., and Ehret, G. (2005). Corticofugal shaping of frequency tuning curves in the central nucleus of the inferior colliculus of mice. *J. Neurophysiol.* 93, 71–83. doi: 10.1152/jn.00348.2004

Yan, W., and Suga, N. (1998). Corticofugal modulation of the midbrain frequency map in the bat auditory system. *Nat. Neurosci.* 1, 54–58.

Yi, H. G., Xie, Z., Reetzke, R., Dimakis, A. G., and Chandrasekaran, B. (2017). Vowel decoding from single-trial speech-evoked electrophysiological responses: a feature-based machine learning approach. *Brain Behav.* 7:e00665. doi: 10.1002/brb3.665

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Cheng, Xu, Gold and Smith. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 22 March 2022
doi: 10.3389/fnins.2022.751595

Structural Connectivity of Human Inferior Colliculus Subdivisions Using *in vivo* and *post mortem* Diffusion MRI Tractography

Kevin R. Sitek[1], Evan Calabrese[2], G. Allan Johnson[3], Satrajit S. Ghosh[4,5] and Bharath Chandrasekaran[1]**

[1] SoundBrain Lab, Brain and Auditory Sciences Research Initiative, Department of Communication and Science Disorders, University of Pittsburgh, Pittsburgh, PA, United States, [2] Department of Radiology and Biomedical Imaging, University of California, San Francisco, San Francisco, CA, United States, [3] Center for In Vivo Microscopy, Duke University, Durham, NC, United States, [4] McGovern Institute for Brain Research, Massachusetts Institute of Technology, Cambridge, MA, United States, [5] Department of Otolaryngology – Head and Neck Surgery, Harvard Medical School, Boston, MA, United States

Inferior colliculus (IC) is an obligatory station along the ascending auditory pathway that also has a high degree of top-down convergence *via* efferent pathways, making it a major computational hub. Animal models have attributed critical roles for the IC in in mediating auditory plasticity, egocentric selection, and noise exclusion. IC contains multiple functionally distinct subdivisions. These include a central nucleus that predominantly receives ascending inputs and external and dorsal nuclei that receive more heterogeneous inputs, including descending and multisensory connections. Subdivisions of human IC have been challenging to identify and quantify using standard brain imaging techniques such as MRI, and the connectivity of each of these subnuclei has not been identified in the human brain. In this study, we estimated the connectivity of human IC subdivisions with diffusion MRI (dMRI) tractography, using both anatomical-based seed analysis as well as unsupervised *k*-means clustering. We demonstrate sensitivity of tractography to overall IC connections in both high resolution *post mortem* and *in vivo* datasets. *k*-Means clustering of the IC streamlines in both the post mortem and *in vivo* datasets generally segregated streamlines based on their terminus beyond IC, such as brainstem, thalamus, or contralateral IC. Using fine-grained anatomical segmentations of the major IC subdivisions, the *post mortem* dataset exhibited unique connectivity patterns from each subdivision, including commissural connections through dorsal IC and lateral lemniscal connections to central and external IC. The subdivisions were less distinct in the context of *in vivo* connectivity, although lateral lemniscal connections were again highest to central and external IC. Overall, the unsupervised and anatomically driven methods provide converging evidence for distinct connectivity profiles for each of the IC subdivisions in both *post mortem* and *in vivo* datasets, suggesting that dMRI tractography with high quality data is sensitive to neural pathways involved in auditory processing as well as top-down control of incoming auditory information.

Keywords: inferior colliculus, diffusion MRI (dMRI), tractography, structural connectivity, human auditory brainstem, subcortical auditory pathway

OPEN ACCESS

Edited by:
Marc Schönwiesner, Leipzig University, Germany

Reviewed by:
David R. Moore, Cincinnati Children's Hospital Medical Center, United States
Robert J. Zatorre, McGill University, Canada

****Correspondence:***
Kevin R. Sitek
kevin.sitek@pitt.edu
Bharath Chandrasekaran
b.chandra@pitt.edu

Specialty section:
This article was submitted to Auditory Cognitive Neuroscience, a section of the journal Frontiers in Neuroscience

Received: *01 August 2021*
Accepted: *27 January 2022*
Published: *22 March 2022*

Citation:
Sitek KR, Calabrese E, Johnson GA, Ghosh SS and Chandrasekaran B (2022) Structural Connectivity of Human Inferior Colliculus Subdivisions Using in vivo and post mortem Diffusion MRI Tractography. Front. Neurosci. 16:751595. doi: 10.3389/fnins.2022.751595

INTRODUCTION

Inferior colliculus (IC) in the dorsal midbrain is a key subcortical auditory structure (Aitkin and Phillips, 1984; Moore, 1987). Evidence from animal models suggests that IC is not only a major computational hub for ascending auditory inputs—perhaps comparable to V1 in the visual system (Nelken et al., 2003; King and Nelken, 2009)—but also a recipient of top-down signals from auditory cortex and thalamus as well as other brain regions implicated in multisensory and cognitive processing (Casseday et al., 2002; Gruters and Groh, 2012).

In animal models, the varying functional roles of IC are largely delineated by major anatomical subdivisions within IC. The central nucleus of the IC (ICc) receives the majority of ascending auditory inputs from the brainstem auditory structures, which arrive *via* the lateral lemniscus (Aitkin and Phillips, 1984; Moore, 1987; Winer et al., 1998; Markovitz et al., 2013; Ono and Ito, 2015). This lemniscal pathway contrasts with non-lemniscal pathways, which largely pass through the dorsal (ICd) and external (ICx) subdivisions of IC. These structures largely receive non-primary ascending auditory inputs as well as lateral and top-down inputs from other brain regions (Suga et al., 1997; Winer et al., 1998; Gruters and Groh, 2012; Straka et al., 2015; Carbajal and Malmierca, 2018; Ito and Malmierca, 2018; Suga, 2020). However, despite their unique functional roles, the major IC subdivisions share wide-ranging connections with both auditory and non-auditory structures (Aitkin, 1989; Oliver, 2005; Winer, 2005; Cant and Oliver, 2018).

While the animal literature on IC subdivision function and connectivity has been built over decades, research in humans has been limited due to the technical challenges of imaging small structures deep within the living human brain. Previous work has demonstrated the feasibility of functional localization and structural connectivity of human subcortical auditory structures *in vivo* (Devlin et al., 2006; Sitek et al., 2019), as well as fundamental sound response properties (Hawley et al., 2005; Sigalovsky and Melcher, 2006; De Martino et al., 2013; Ress and Chandrasekaran, 2013; Moerel et al., 2015). Recent work has investigated functional subdivisions of human auditory thalamus (Mihai et al., 2019; Tabas et al., 2020). However, to our knowledge no work has investigated subdivisions of human IC in living humans or their patterns of connectivity beyond IC.

Due to the lack of clarity in the literature regarding human IC subdivision connectivity, we sought to identify the white matter connectivity of human IC subdivisions using three high quality datasets to establish the feasibility of human IC subdivision connectivity measurements with diffusion-weighted MRI tractography. Using a 200 μm isotropic resolution *post mortem* sample, a high resolution 760 μm isotropic diffusion MRI (dMRI) dataset from a single living human participant, and a near-millimeter resolution 10-subject 7T dMRI dataset, we used *k*-means clustering to identify unique connectivity patterns of human IC. We then identified the major subdivisions of IC and estimated tractography from each subdivision. Our findings suggest that dMRI tractography is sensitive to fine-grained structural connectivity within human IC.

MATERIALS AND METHODS

MRI Data Acquisition

Three unique datasets were included in this study. The first is a *post mortem* human brainstem of a 65-year-old male who died of non-neurological natural causes (Calabrese et al., 2015; Sitek et al., 2019; Rushmore et al., 2020; Adil et al., 2021). The tissue was removed approximately 24 h *post mortem* and fixed with 10% formalin solution for 2 weeks. The tissue was rehydrated in saline with 1% gadoteridol 1 week before MRI acquisition. For imaging purposes, the tissue was placed in a fluorocarbon liquid in a custom MRI-compatible tube. DMRI was collected in a small-bore 7-Tesla MRI over 208 h at b = 4,000 s/mm^2 in 120 diffusion directions at 200 μm isotropic resolution. Anatomical T2*-weighted images were collected at 50 μm isotropic resolution.

The second dataset is a single *in vivo* participant (about 30 years old) scanned in a 3-Tesla Siemens Connectom scanner over 18 h at b = 1,000 and 2,500 s/mm^2 in 1,260 diffusion directions at 760 μm isotropic resolution (see Wang et al., 2021 for complete acquisition details). The participant gave written informed consent for participation in the study, which was approved by the Institutional Review Board of Partners Healthcare.

Finally, we used an existing *in vivo* dataset of 10 individuals (25–30 years of age) scanned in a 7-Tesla Siemens Magnetom MRI at 1.05 mm isotropic resolution (Sitek et al., 2019). The dMRI acquisition was based on the 7T Human Connectome Project acquisition (Vu et al., 2015) and extended from two shells to three (b = 1,000, 2,000, 3,000 s/mm^2) (Gulban et al., 2018). This experiment was approved by the ethics committee of Maastricht University (protocol number ERCPN-167_09_05_2016), and each participant provided written informed consent for participation in the study.

Diffusion MRI Processing

Each volume of the *post mortem* dMRI data was affine-transformed to the first b0 image volume using ANTs tools (Avants et al., 2011) in order to correct for eddy current distortions (Calabrese et al., 2015). The *post mortem* dMRI data were then linearly transformed to the T2*-weighted anatomical MRI (Sitek et al., 2019).

The sub-millimeter *in vivo* images were corrected for susceptibility-induced distortion, eddy-current distortion, gradient non-linearity, and subject motion using FSL tools (Jenkinson et al., 2012) as described in Wang et al. (2021).

Diffusion orientation estimation for the *post mortem* and sub-millimeter *in vivo* MRI was performed in DSI Studio using generalized q-sampling imaging (GQI) (Yeh et al., 2010), including the generation of quantitative anisotropy maps [similar to fractional anisotropy (FA) maps generated in diffusion tensor imaging (DTI) analysis].

Post mortem dataset processing was performed locally on an Intel-based Macbook Pro using the 7 January 2021 build of DSI Studio; the *in vivo* sub-millimeter dataset was processed in DSI Studio by F. C. Yeh and shared publicly at https://brain.labsolver.org/mgh_760.html.

The 7T *in vivo* dataset was processed using the HCP pipeline (Sotiropoulos et al., 2013; Glasser et al., 2016), including geometric and eddy-current distortion correction and motion compensation. Data were masked to only include the brainstem and thalamus, with diffusion fiber orientation distributions estimated using constrained spherical convolution implemented in DIPY (Tournier et al., 2004; Garyfallidis et al., 2014). Streamlines were generated using DIPY's EuDX approach.

Data Analysis

We conducted two analyses in each dataset. In the unsupervised *k*-means clustering analysis, the entire IC was used as a tractography seed in DSI Studio (Yeh et al., 2010). The resulting streamlines were partitioned using *k*-means clustering, run with multiple values of *k* ranging from 2 to 10. In the seed-based analysis, IC subdivisions were manually labeled on the anatomical images, and tractography was run from each subdivision.

In both analyses, the anisotropy threshold was randomly selected between 0.5 and 0.7 times the Otsu threshold (maximizing contrast between foreground and background), the angular threshold was 90°, and the step size was randomly selected from 0.5 to 1.5 voxels. Each individual tractography operation used 10 million seed points with subvoxel seeding and utilizing all fiber orientations in each voxel.

Segmenting Inferior Colliculus Subdivisions

We segmented IC following human histological literature utilizing a variety of staining techniques, including Nissl, myelin, and AChE (Moore, 1987; Webster, 1992; Mansour et al., 2019; Paxinos et al., 2020). Although there are slight differences in terminology throughout the literature, we divided IC into three major subdivisions: central nucleus of the inferior colliculus (ICc), external cortex of the IC (ICx), and dorsal cortex of the inferior colliculus (ICd).

In the *post mortem* sample, overall IC segmentations were originally generated in an atlas of human subcortical structures (Sitek et al., 2019) and are publicly available at https://osf.io/c4m82/. For this investigation, we used the conjunction atlas (based on two raters) that was then dilated 500 μm (about 2.5 voxels)—see section "Discussion" for further discussion. Within IC, ICc was identified as the large, moderate intensity (in the T2*-weighted image) area near the center of each IC (**Figure 1**). ICx appeared as a dark band lateral to ICc in the T2*-weighted images. ICd was segmented as the hyperintense zone rostral to ICc.

In the *in vivo* datasets, IC was identified on axial views of the dMRI quantitative anisotropy map as caudal structures of the dorsal tectum, bounded by PAG ventrally, superior colliculus rostrally, and CSF dorsally and laterally (**Figure 2**). Within IC, ICc was visible on the quantitative anisotropy images a dark structure at the center of IC. Meanwhile, ICx could be found as a lighter band lateral to ICc, while ICd was identifiable as a lighter band posterior to ICc.

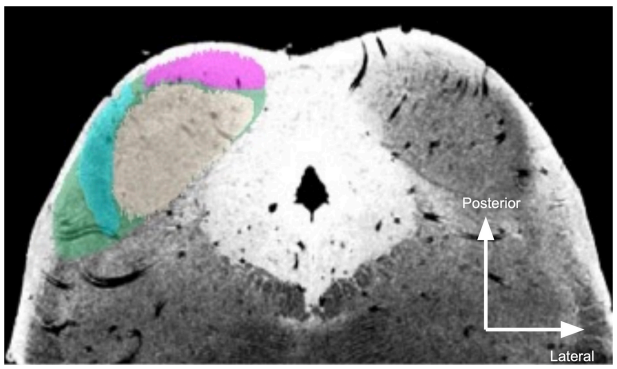

FIGURE 1 | Segmentation of *post mortem* inferior colliculus and its major subdivisions. Neutral gray, central nucleus (ICc); magenta, dorsal nucleus (ICd); turquoise, external nucleus (ICx); green, overall inferior colliculus (IC). The right IC is unsegmented to show MR contrast. Segmentations were performed manually on the 50 μm isotropic T2*-weighted anatomical dataset and transformed to dMRI space (200 μm isotropic).

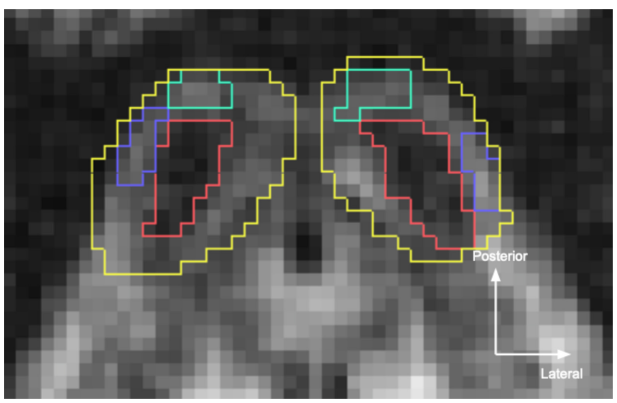

FIGURE 2 | Inferior colliculus (IC) segmentations in the sub-millimeter *in vivo* dataset. Yellow, overall IC; red, central nucleus (ICc); turquoise, dorsal nucleus (ICd); blue, external nucleus (ICx). Segmentations were hand drawn on the 760 μm isotropic diffusion MRI quantitative anisotropy map.

In dMRI tractography, streamlines (sometimes referred to as "tracks" or "tracts") are volumeless representations of likely white matter pathways. To determine the paths of streamlines beyond IC, we segmented three white matter structures adjacent to IC: lateral lemniscus, which connects IC with the more caudal brainstem auditory structures; brachium of the IC, which connects IC with more rostral thalamic and cortical auditory structures; and commissure of the IC, which connects the two colliculi. To compare connectivity patterns between IC subdivisions, we then counted the number of streamlines from each IC subdivision to each white matter structure.

To assess the specificity of *k*-means clustering streamline clusters with respect to anatomically defined IC subdivisions, in the 7T *in vivo* dataset, we assessed the number of streamlines per *k*-means cluster (with the number of clusters ranging from $k = 2–10$) that passed through each IC subdivision. For each IC,

k, and cluster with sufficient streamlines, we computed an ROI–streamline FA score, using the number of streamlines passing through each IC subdivision. This provides a quantitative value (from 0 to 1) that represents the a given streamline cluster's specificity for each IC subdivision and allows us to discern whether a particular number of *k*-means clusters is optimally aligned with anatomically segregated streamlines.

However, there is debate about the utility of streamline counts, which vary based on acquisition and analysis details (Smith et al., 2012, 2015, 2020; Yeh et al., 2019; Rheault et al., 2020); specific values should be interpreted with caution.

RESULTS

Diffusion MRI Tractography of Inferior Colliculus

We first ran tractography using the entire IC as a seed ROI. Starting with 10 million seed points, in the *post mortem* dataset, we generated 3,707 streamlines from left IC and 5,143 streamlines from right IC (**Figure 3**, left). In the sub-millimeter *in vivo* dataset, we generated 629 streamlines from left IC and 1,017 streamlines from right IC (**Figure 3**, right). In the 7T *in vivo* dataset (which was preprocessed separately and modeled using constrained spherical deconvolution as opposed to generalized GQI like the *post mortem* and sub-millimeter *in vivo* datasets), averaged across the 10 participants, we generated 95.4 left IC streamlines and 98.4 right IC streamlines. Overall, from visual inspection of the generated dMRI results (**Figure 3**), streamlines that pass through IC run caudally through lateral lemniscus to the brainstem auditory structures, rostrally through the brachium of the IC to the thalamic (medial geniculate) and cortical auditory structures, and laterally through the commissure of the IC to the contralateral IC. In addition, we identified streamlines passing through non-primary auditory structures, such as superior colliculus (rostral/superior relative to IC) and cerebellar peduncles (caudal/inferior and posterior relative to IC; **Figure 3**).

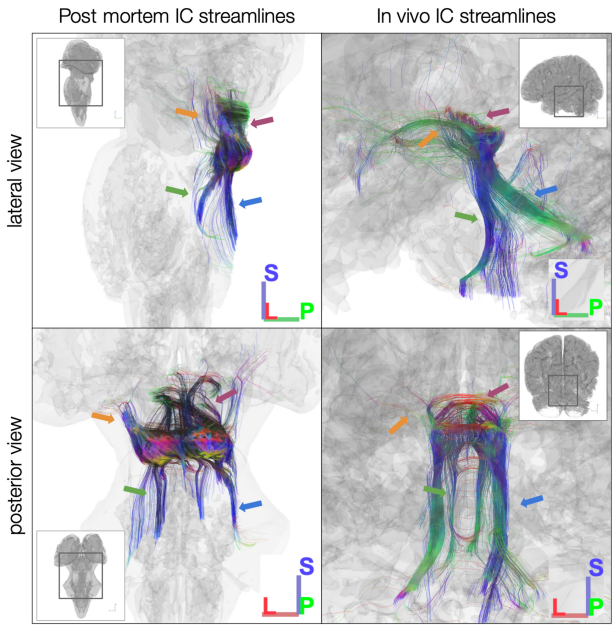

FIGURE 3 | Diffusion MRI tractography streamlines passing through left and right inferior colliculus (IC). Colors represent local streamline orientation: blue, superior–inferior; red, left–right; green, anterior–posterior. Insets show entire specimen or brain surface. Left: *post mortem* dataset. Right: *in vivo* MGH 760 μm dataset. Both datasets demonstrate canonical auditory pathways (green arrows, lateral lemniscus; orange arrows, brachium of IC) as well as non-primary auditory connections [for instance, to superior colliculus (magenta arrows) and through cerebellar peduncles (blue arrows)].

k-Means Clustering of Inferior Colliculus Streamlines

In the *post mortem* dataset, *k*-means clustering isolated streamlines rostrally between IC and thalamus/cortex and caudally between IC and brainstem (**Figure 4**). Although clustering was performed separately for each IC, the three resulting clusters were similar across left and right ICs. For instance, Cluster 2 in each IC consisted largely of caudal streamlines, with limited rostral streamlines and very few commissural streamlines (**Figure 4**, right). Meanwhile, Clusters 1 and 3 exhibited more commissural streamlines in both ICs.

In the sub-millimeter *in vivo* dataset, the reduced number of streamlines limit the interpretability of the *k*-means clustering results, although the clusters in left and right IC have similar connectivity patterns despite being generated in separate *k*-means clustering operations (**Figure 5**). In left and right IC, Cluster 1 contains primarily caudal-extending streamlines that also reach toward the midline at the level of the IC or SC. Cluster 2 in both ICs captured many of the cerebellar streamlines, while Cluster 3 has predominantly rostral-extending streamlines that also extend caudally toward brainstem and cerebellum.

Anatomically Defined Inferior Colliculus Subdivision Connectivity

Using anatomically defined IC subdivisions as tractography seeds, streamlines were less extensive than when using the whole IC segmentations as tractography seeds. In the *post mortem* dataset, the central nucleus of the IC (ICc) had the fewest streamlines extending rostrally toward the medial geniculate body (MGB) of the thalamus (**Figure 6**). Meanwhile, the dorsal nucleus of the IC (ICd) had by far the most streamlines crossing the midline to the contralateral IC. Additionally, the external nucleus of the IC (ICx) exhibited more streamlines extending rostrally toward MGB than caudally toward brainstem.

The sub-millimeter *in vivo* dataset (**Figure 7**) demonstrated sparser connectivity than the *post mortem* dataset to IC subdivisions. Unlike the *post mortem* dataset, few streamlines from any subdivision extended rostrally toward thalamus and cortex or crossed the midline to the contralateral IC. Caudal-extending streamlines were more frequent, particularly from ICc and ICx on both the left and right.

To assess overall connectivity patterns, we counted the number of streamlines from each IC subdivision that reached

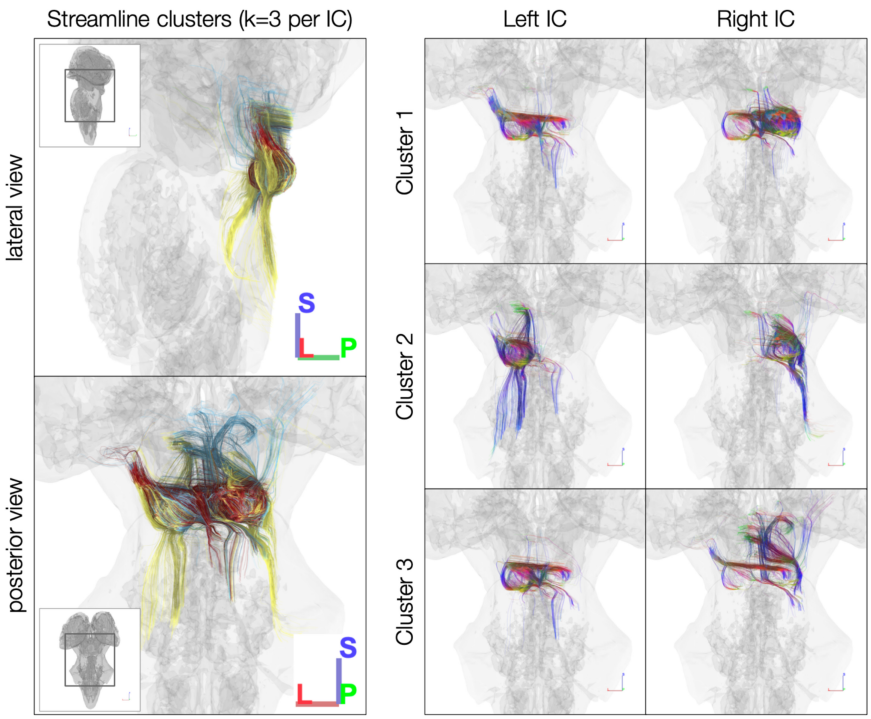

FIGURE 4 | *k*-Means clustering of inferior colliculus (IC) streamlines (k = 3 per IC). Left: color denotes cluster (1, 2, or 3, based on *k*-means clustering). Right: color denotes local streamline orientation: blue, superior–inferior; red, left–right; green, anterior–posterior.

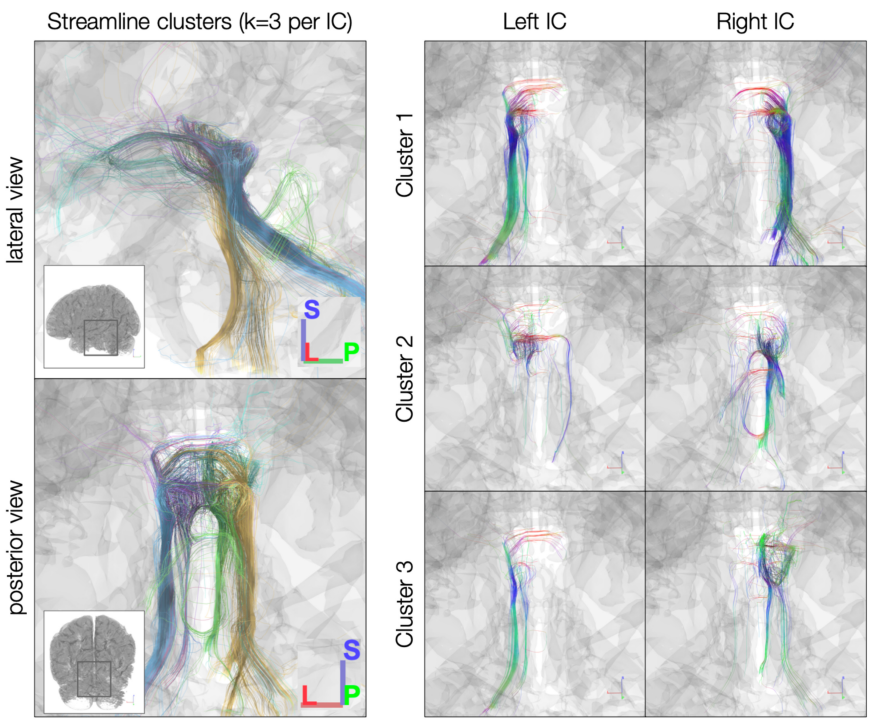

FIGURE 5 | *k*-Means clustering [k = 3 per inferior colliculus (IC)] of the *in vivo* 760 μm dataset. Left: color denotes cluster (1, 2, or 3, based on *k*-means clustering). Right, color denotes local streamline orientation; blue, superior–inferior; red, left–right; green, anterior–posterior.

each white matter structure (lateral lemniscus, brachium of the IC, and commissure of the IC). Of the white matter structures, only the IC commissure showed a preference for streamlines from any particular IC substructure in the *post mortem* dataset, with a large proportion of commissural streamlines passing through ICd contralaterally (**Figure 8**).

Due to the reduced number of overall IC streamlines in the sub-millimeter *in vivo* dataset, we found fewer streamlines from IC subdivisions reaching the white matter structures (**Figure 9**). Right ICd contributed the most streamlines to commissural connections, but interestingly not left ICd. Compared to the *post mortem* dataset, a higher proportion of lateral lemniscal streamlines passed through IC substructures, particularly left and right ICx (as well as right ICc).

In the 10-participant 7T *in vivo* dataset, we again saw strong commissural connections through dorsal IC (**Figure 10**). In this dataset, the brachium of the IC and the lateral lemniscus shared connectivity patterns, with similar streamline proportions reaching central and external IC nuclei.

Relating *k*-Means and Anatomical Approaches

To assess whether a particular number of *k*-means clusters is optimal for segregating subdivision streamlines, we ran *k*-means clustering with *k* varying from 2 to 10 on streamlines passing through each IC in the 10-subject *in vivo* dataset. We counted the number of streamlines in each cluster passing through each anatomically defined IC subdivision and, to determine whether a given cluster had a specific subdivisional correspondence, calculated a "ROI–subdivision fractional anisotropy" score for each *k*-value and IC. Upon inspection, many of the clusters at higher *k*-values had few or no streamlines, due to the small number of streamlines to begin with. We therefore focused our investigation on $k = 2–5$. Across *k*-values from 2 to 5, the mean score across subjects ranged from 0.61 to 0.69. We did not find a significant difference between ROI–subdivision FA scores from different *k*-values (Kruskal–Wallis $h = 0.50$, $p = 0.92$), suggesting there was not an optimal number of clusters that aligned best with subdivision anatomy.

DISCUSSION

Using sub-millimeter and near-millimeter resolution dMRI from high quality *post mortem* and *in vivo* human datasets, we investigated substructure connectivity patterns of human IC. As gold standard methods such as tracer injections are challenging or impossible with human tissue, dMRI tractography represents the best opportunity to map fine-grained connections in the human brain (Barbeau et al., 2020). With unsupervised *k*-means clustering approaches to cluster white matter connections through IC, we found that streamline clusters were segregated based on their origin beyond IC as well as their location within IC. For instance, in both the *post mortem* and *in vivo* datasets, caudal-extending streamlines (toward brainstem through the lateral lemniscus) were largely separated into their own cluster, reaching the IC in distinct locations. The results aligned with connectivity patterns based on tractography analysis using anatomically defined IC subdivisions: both *post mortem* and *in vivo* datasets demonstrated lateral lemniscal streamlines through central and external IC nuclei, as well as strong commissural connections through dorsal IC nuclei. Taken together, the *k*-means clustering and anatomically driven analyses demonstrate that dMRI tractography can reveal fine-grained structural connectivity patterns within the human subcortical auditory system. By utilizing diverse state-of-the-art datasets with results largely in agreement, our results help build consensus around the utility of dMRI tractography for investigating subcortical auditory connectivity.

Although many of the streamlines we observed in tractography results align with the major auditory pathways (including lateral lemniscus and brachium of the IC), we also identified streamlines heading toward non-auditory structures (**Figure 3**). For instance, in both our *post mortem* and *in vivo* datasets, superior colliculus received many streamlines that were generated with a tractography seed place in the external nucleus of IC (ICx), whose connections to superior colliculus have been described previously in animal models (Edwards et al., 1979; Druga and Syka, 1984; Doubell et al., 2000; Bednárová et al., 2018). Similarly, both datasets showed IC streamlines running toward cerebellum (Powell and Hatton, 1969). As we continue mapping and quantifying the structural connectivity of human IC, it is important to keep in mind that auditory processing plays a critical role in motor, limbic, and multisensory processing, and the connections with non-primary auditory structures that have been previously identified in animal models may be crucial infrastructure for these complex neural functions.

This work builds on previous literature examining the structural connectivity of human IC as a whole. Some groups have used probabilistic tractography to estimate connectivity between other auditory structures and IC (Devlin et al., 2006; Javad et al., 2014). Others used IC as a landmark for estimating subcortical auditory connectivity in clinical populations (Lin et al., 2008; Wu et al., 2009; Tarabichi et al., 2018). Our own previous work established reliable *post mortem* and *in vivo* tractography estimates of connectivity throughout the subcortical auditory system, including to IC (Sitek et al., 2019). However, to our knowledge, no previous work has investigated the connectivity patterns of IC's constituent nuclei.

Much of this difficulty arises from accurately segmenting the anatomical boundaries of IC subdivisions in living humans. Indeed, we did not find any previous MRI investigations of *in vivo* human IC subdivisions, and *post mortem* MRI atlases have been varying in detailing IC substructure (Paxinos et al., 2012). In the present work, we take advantage of the high MR contrast and ultra-high resolution *post mortem* MRI to finely delineate the major subdivisions of human IC, which we then used as the basis for segmenting the IC subdivisions in the included *in vivo* dataset.

Despite the advances of the present work, applying the methods to standard *in vivo* diffusion-weighted MRI pose significant challenges. Each of the datasets used in this study was acquired over multiple days in unique MRI environments with specialized scanning protocols. In contrast, diffusion-weighted MRI is typically collected in a single session—often in

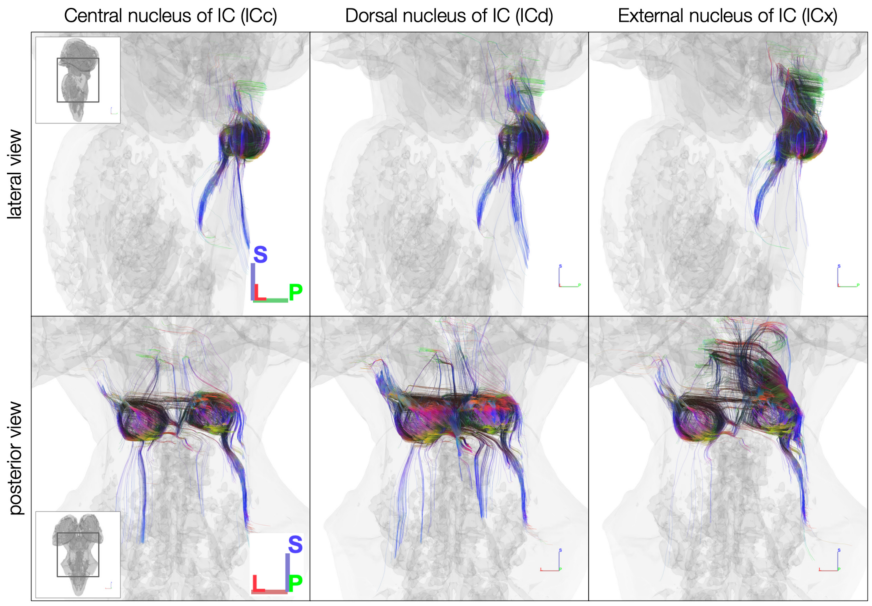

FIGURE 6 | Diffusion MRI tractography streamlines reaching each of the inferior colliculus subdivisions in the *post mortem* dataset. Colors represent local streamline orientation: blue, superior–inferior; red, left–right; green, anterior–posterior.

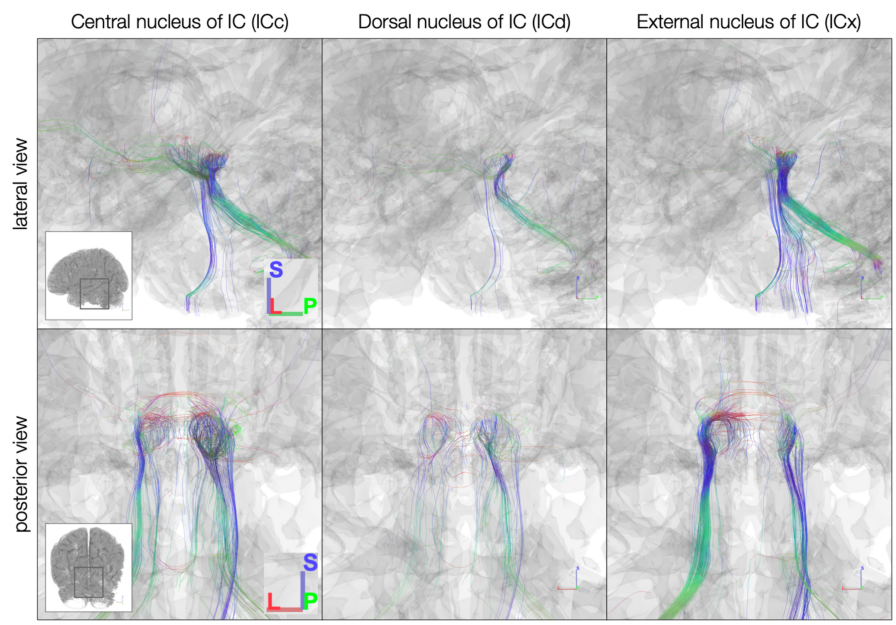

FIGURE 7 | *In vivo* anatomically defined inferior colliculus (IC) subdivision streamlines. Colors represent local streamline orientation: blue, superior–inferior; red, left–right; green, anterior–posterior.

just one or two 5–10-min acquisitions—on standard 3T MRI scanners, which limits the potential spatial resolution, angular resolution, diffusion sensitivity, and contrast-to-noise ratio of the collected images (McNab et al., 2013). Further, there remain outstanding issues in the implementation and interpretation of dMRI tractography, such as a lack of specificity resulting in many false positives (Thomas et al., 2014; Schilling et al., 2019). Additionally, there are unresolved issues when quantifying streamlines and connections (Jbabdi and Johansen-Berg, 2011; Jeurissen et al., 2019; Smith et al., 2020), which limits the interpretation of specific streamline counts in the present study. However, advances in dMRI acquisition and analysis, including at ultra-high magnetic fields, are improving the sensitivity and reliability of dMRI tractography (Setsompop et al., 2013;

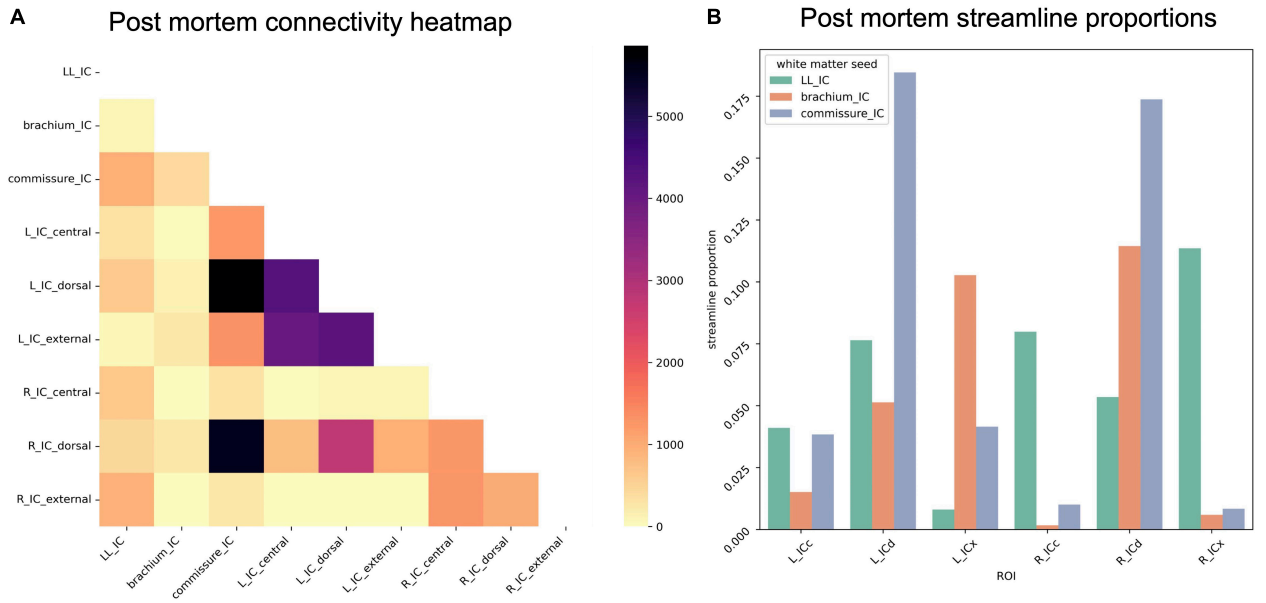

FIGURE 8 | (A) Connectivity matrix of *post mortem* inferior colliculus (IC) subdivisions and adjacent white matter tracts (values indicate number of streamlines); **(B)** proportion of streamlines reaching each white matter region for each IC subdivision.

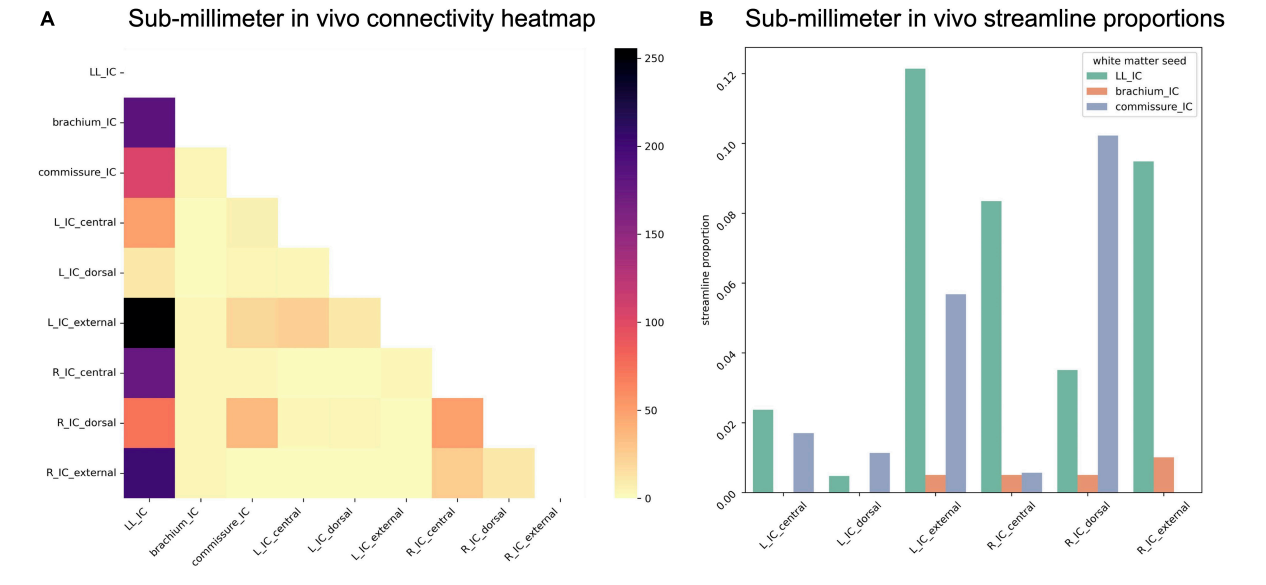

FIGURE 9 | (A) Connectivity matrix of sub-millimeter *in vivo* inferior colliculus (IC) subdivisions and adjacent white matter tracts (values indicate number of streamlines); **(B)** proportion of streamlines reaching each white matter region for each IC subdivision.

Sotiropoulos et al., 2013; Vu et al., 2015; Jeurissen et al., 2019; Kruper et al., 2021; Moeller et al., 2021; Yeh et al., 2021), making finer-grained connectivity investigations more accessible to the broader neuroimaging community. Additional *post mortem* human dMRI datasets (Edlow et al., 2019; Tendler et al., 2021)—along with complementary cellular-level resolution methods such as polarized light imaging (Axer et al., 2011) and polarization-sensitive optical coherence tomography (Jones et al., 2020)—will provide critical details on human brainstem 3-D anatomy.

In general, sub-millimeter resolution dMRI is necessary for brainstem tractography in order to dissociate densely packed nuclei and white matter pathways (Ford et al., 2013; Grisot et al., 2021; Yendiki et al., 2021). However, imaging at high resolution may introduce its own challenges in identifying connectivity between nuclei, particularly in brainstem and other non-cortical brain structures. For example, as we have previously discussed (Sitek et al., 2019), the improved contrast and spatial specificity between gray matter and white matter results in decreased

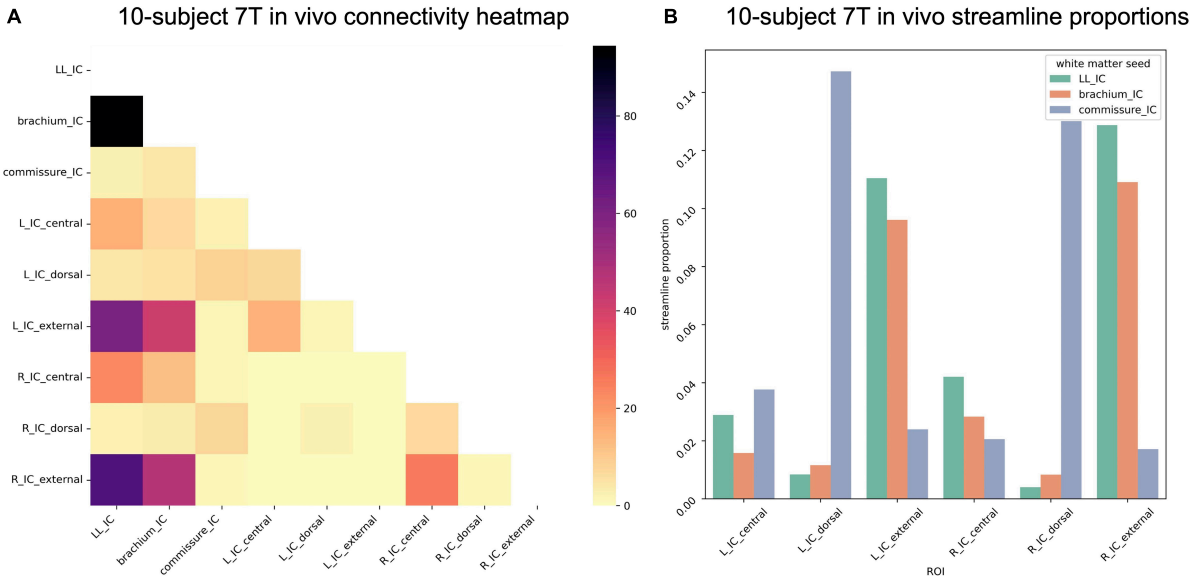

FIGURE 10 | (A) Connectivity matrix of 7T *in vivo* inferior colliculus (IC) subdivisions and adjacent white matter tracts (values indicate number of streamlines averaged across 10 participants); **(B)** proportion of streamlines reaching each white matter region for each IC subdivision.

partial volume effects, with a (unsurprising, but perhaps overlooked) result of streamlines staying within the white matter and fewer streamlines reaching finely segmented gray matter structures. This may have affected our present results, where the anatomically defined subdivision segmentations generally excluded adjacent white matter and thus may not demonstrate the full connectivity patterns of these regions. For this reason, we opted to use the 500 μm-dilated whole IC segmentation from Sitek et al. (2019), which demonstrated improved connectivity profiles relative to the strict IC gray matter segmentation.

However, with the development of MRI hardware capable of stronger diffusion encoding, as well as continued research into optimal preprocessing and analysis methods, we are hopeful that IC subdivision tractography will yield new insights into the relationship between subcortical auditory connectivity and perception-dependent human behavior such as speech communication and music, as well as the role of human IC subdivisions in health and disease.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://github.com/SoundBrainLab/IC_subdivision_connectivity.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Institutional Review Board of Partners Healthcare and the Ethics Committee of the Faculty of Psychology and Neuroscience of the University of Maastricht. The patients/participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

KS analyzed the data, wrote the manuscript, and submitted the manuscript. EC and GJ collected and preprocessed the *post mortem* data and provided feedback on the manuscript. SG provided guidance on analysis and feedback on the manuscript. BC provided guidance on analysis, contributed to and provided feedback on the manuscript, and provided funding for the research. All authors contributed to the article and approved the submitted version.

FUNDING

KS and BC were funded by R01-DC013315 and R01-DC015504 (both awarded to BC). SG was partially supported by P41-EB019936. GJ was partially supported by P41-EB015897 and NIH 1S10OD010683.

ACKNOWLEDGMENTS

We thank Fuyixue Wang and colleagues for collecting and sharing the sub-millimeter *in vivo* dataset, we thank Fang-Cheng Yeh for analyzing and sharing the dataset, and we also thank Omer Faruk Gulban for collecting and sharing the 7T *in vivo* dataset.

REFERENCES

Adil, S. M., Calabrese, E., Charalambous, L. T., Cook, J. J., Rahimpour, S., Atik, A. F., et al. (2021). A high-resolution interactive atlas of the human brainstem using magnetic resonance imaging. *Neuroimage* 237:118135. doi: 10.1016/j.neuroimage.2021.118135

Aitkin, L. (1989). "Auditory system," in *Handbook of Chemical Neuroanatomy, Volume 7: Integrated Systems of the CNS, Part II: Central Visual, Auditory, Somatosensory, Gustatory*, eds A. Björklund, T. Hökfelt, and L. W. Swanson (Leningrad: Nauka).

Aitkin, L. M., and Phillips, S. C. (1984). Is the inferior colliculus an obligatory relay in the cat auditory system? *Neurosci. Lett.* 44, 259–264. doi: 10.1016/0304-3940(84)90032-6

Avants, B. B., Tustison, N. J., Song, G., Cook, P. A., Klein, A., and Gee, J. C. (2011). A reproducible evaluation of ANTs similarity metric performance in brain image registration. *Neuroimage* 54, 2033–2044. doi: 10.1016/j.neuroimage.2010.09.025

Axer, H., Beck, S., Axer, M., Schuchardt, F., Heepe, J., Flücken, A., et al. (2011). Microstructural analysis of human white matter architecture using polarized light imaging: views from neuroanatomy. *Front. Neuroinform.* 5:28. doi: 10.3389/fninf.2011.00028

Barbeau, E. B., Descoteaux, M., and Petrides, M. (2020). Dissociating the white matter tracts connecting the temporo-parietal cortical region with frontal cortex using diffusion tractography. *Sci. Rep.* 10:8186. doi: 10.1038/s41598-020-64124-y

Bednárová, V., Grothe, B., and Myoga, M. H. (2018). Complex and spatially segregated auditory inputs of the mouse superior colliculus. *J. Physiol. (Lond.)* 596, 5281–5298. doi: 10.1113/JP276370

Calabrese, E., Hickey, P., Hulette, C., Zhang, J., Parente, B., Lad, S. P., et al. (2015). Postmortem diffusion MRI of the human brainstem and thalamus for deep brain stimulator electrode localization. *Hum. Brain Mapp.* 36, 3167–3178. doi: 10.1002/hbm.22836

Cant, N. B., and Oliver, D. L. (2018). "Overview of auditory projection pathways and intrinsic microcircuits," in *The Mammalian Auditory Pathways Springer Handbook of Auditory Research*, eds D. L. Oliver, N. B. Cant, R. R. Fay, and A. N. Popper (Cham: Springer International Publishing), 7–39.

Carbajal, G. V., and Malmierca, M. S. (2018). The neuronal basis of predictive coding along the auditory pathway: from the subcortical roots to cortical deviance detection. *Trends Hear.* 22:2331216518784822. doi: 10.1177/2331216518784822

Casseday, J. H., Fremouw, T., and Covey, E. (2002). "The inferior colliculus: a hub for the central auditory system," in *Integrative Functions in the Mammalian Auditory Pathway Springer Handbook of Auditory Research*, eds D. Oertel, R. R. Fay, and A. N. Popper (New York, NY: Springer New York), 238–318.

De Martino, F., Moerel, M., van de Moortele, P.-F., Ugurbil, K., Goebel, R., Yacoub, E., et al. (2013). Spatial organization of frequency preference and selectivity in the human inferior colliculus. *Nat. Commun.* 4:1386. doi: 10.1038/ncomms2379

Devlin, J. T., Sillery, E. L., Hall, D. A., Hobden, P., Behrens, T. E. J., Nunes, R. G., et al. (2006). Reliable identification of the auditory thalamus using multi-modal structural analyses. *Neuroimage* 30, 1112–1120. doi: 10.1016/j.neuroimage.2005.11.025

Doubell, T. P., Baron, J., Skaliora, I., and King, A. J. (2000). Topographical projection from the superior colliculus to the nucleus of the brachium of the inferior colliculus in the ferret: convergence of visual and auditory information. *Eur. J. Neurosci.* 12, 4290–4308. doi: 10.1111/j.1460-9568.2000.01337.x

Druga, R., and Syka, J. (1984). Projections from auditory structures to the superior colliculus in the rat. *Neurosci. Lett.* 45, 247–252. doi: 10.1016/0304-3940(84)90234-9

Edlow, B. L., Mareyam, A., Horn, A., Polimeni, J. R., Witzel, T., Tisdall, M. D., et al. (2019). 7 Tesla MRI of the ex vivo human brain at 100 micron resolution. *Sci. Data* 6:244. doi: 10.1038/s41597-019-0254-8

Edwards, S. B., Ginsburgh, C. L., Henkel, C. K., and Stein, B. E. (1979). Sources of subcortical projections to the superior colliculus in the cat. *J. Comp. Neurol.* 184, 309–329. doi: 10.1002/cne.901840207

Ford, A. A., Colon-Perez, L., Triplett, W. T., Gullett, J. M., Mareci, T. H., and Fitzgerald, D. B. (2013). Imaging white matter in human brainstem. *Front. Hum. Neurosci.* 7:400. doi: 10.3389/fnhum.2013.00400

Garyfallidis, E., Brett, M., Amirbekian, B., Rokem, A., van der Walt, S., Descoteaux, M., et al. (2014). Dipy, a library for the analysis of diffusion MRI data. *Front. Neuroinform.* 8:8. doi: 10.3389/fninf.2014.00008

Glasser, M. F., Smith, S. M., Marcus, D. S., Andersson, J. L. R., Auerbach, E. J., Behrens, T. E. J., et al. (2016). The human connectome Project's neuroimaging approach. *Nat. Neurosci.* 19, 1175–1187. doi: 10.1038/nn.4361

Grisot, G., Haber, S. N., and Yendiki, A. (2021). Diffusion MRI and anatomic tracing in the same brain reveal common failure modes of tractography. *Neuroimage* 239:118300. doi: 10.1016/j.neuroimage.2021.118300

Gruters, K. G., and Groh, J. M. (2012). Sounds and beyond: multisensory and other non-auditory signals in the inferior colliculus. *Front. Neural Circuits* 6:96. doi: 10.3389/fncir.2012.00096

Gulban, O. F., De Martino, F., Vu, A. T., Yacoub, E., Uğurbil, K., and Lenglet, C. (2018). Cortical fibers orientation mapping using in-vivo whole brain 7 T diffusion MRI. *Neuroimage* 178, 104–118. doi: 10.1016/j.neuroimage.2018.05.010

Hawley, M. L., Melcher, J. R., and Fullerton, B. C. (2005). Effects of sound bandwidth on fMRI activation in human auditory brainstem nuclei. *Hear. Res.* 204, 101–110. doi: 10.1016/j.heares.2005.01.005

Ito, T., and Malmierca, M. S. (2018). "Neurons, connections, and microcircuits of the inferior colliculus," in *The Mammalian Auditory Pathways Springer Handbook of Auditory Research*, eds D. L. Oliver, N. B. Cant, R. R. Fay, and A. N. Popper (Cham: Springer International Publishing), 127–167. doi: 10.1007/978-3-319-71798-2_6

Javad, F., Warren, J. D., Micallef, C., Thornton, J. S., Golay, X., Yousry, T., et al. (2014). Auditory tracts identified with combined fMRI and diffusion tractography. *Neuroimage* 84, 562–574. doi: 10.1016/j.neuroimage.2013.09.007

Jbabdi, S., and Johansen-Berg, H. (2011). Tractography: where do we go from here? *Brain Connect.* 1, 169–183. doi: 10.1089/brain.2011.0033

Jenkinson, M., Beckmann, C. F., Behrens, T. E., Woolrich, M. W., and Smith, S. M. (2012). FSL. *Neuroimage* 62, 782–790. doi: 10.1016/j.neuroimage.2011.09.015

Jeurissen, B., Descoteaux, M., Mori, S., and Leemans, A. (2019). Diffusion MRI fiber tractography of the brain. *NMR Biomed.* 32:e3785. doi: 10.1002/nbm.3785

Jones, R., Grisot, G., Augustinack, J., Magnain, C., Boas, D. A., Fischl, B., et al. (2020). Insight into the fundamental trade-offs of diffusion MRI from polarization-sensitive optical coherence tomography in ex vivo human brain. *Neuroimage* 214:116704. doi: 10.1016/j.neuroimage.2020.116704

King, A. J., and Nelken, I. (2009). Unraveling the principles of auditory cortical processing: can we learn from the visual system? *Nat. Neurosci.* 12, 698–701. doi: 10.1038/nn.2308

Kruper, J., Yeatman, J. D., Richie-Halford, A., Bloom, D., Grotheer, M., Caffarra, S., et al. (2021). Evaluating the reliability of human brain white matter tractometry. *BioRxiv [preprint]* doi: 10.1101/2021.02.24.432740

Lin, Y., Wang, J., Wu, C., Wai, Y., Yu, J., and Ng, S. (2008). Diffusion tensor imaging of the auditory pathway in sensorineural hearing loss: changes in radial diffusivity and diffusion anisotropy. *J. Magn. Reson. Imaging* 28, 598–603. doi: 10.1002/jmri.21464

Mansour, Y., Altaher, W., and Kulesza, R. J. (2019). Characterization of the human central nucleus of the inferior colliculus. *Hear. Res.* 377, 234–246. doi: 10.1016/j.heares.2019.04.004

Markovitz, C. D., Tang, T. T., and Lim, H. H. (2013). Tonotopic and localized pathways from primary auditory cortex to the central nucleus of the inferior colliculus. *Front. Neural Circuits* 7:77. doi: 10.3389/fncir.2013.00077

McNab, J. A., Edlow, B. L., Witzel, T., Huang, S. Y., Bhat, H., Heberlein, K., et al. (2013). The human connectome project and beyond: initial applications of 300 mT/m gradients. *Neuroimage* 80, 234–245. doi: 10.1016/j.neuroimage.2013.05.074

Mihai, P. G., Moerel, M., de Martino, F., Trampel, R., Kiebel, S., and von Kriegstein, K. (2019). Modulation of tonotopic ventral medial geniculate body is behaviorally relevant for speech recognition. *Elife* 8:e44837. doi: 10.7554/eLife.44837

Moeller, S., Pisharady Kumar, P., Andersson, J., Akcakaya, M., Harel, N., Ma, R. E., et al. (2021). Diffusion imaging in the post HCP era. *J. Magn. Reson. Imaging* 54, 36–57. doi: 10.1002/jmri.27247

Moerel, M., De Martino, F., Uğurbil, K., Yacoub, E., and Formisano, E. (2015). Processing of frequency and location in human subcortical auditory structures. *Sci. Rep.* 5:17048. doi: 10.1038/srep17048

Moore, J. K. (1987). The human auditory brain stem: a comparative view. *Hear. Res.* 29, 1–32. doi: 10.1016/0378-5955(87)90202-4

Nelken, I., Fishbach, A., Las, L., Ulanovsky, N., and Farkas, D. (2003). Primary auditory cortex of cats: feature detection or something else? *Biol. Cybern.* 89, 397–406. doi: 10.1007/s00422-003-0445-3

Oliver, D. L. (2005). "Neuronal organization in the inferior colliculus," in *The Inferior Colliculus*, eds J. A. Winer and C. E. Schreiner (New York, NY: Springer-Verlag), 69–114. doi: 10.1007/0-387-27083-3_2

Ono, M., and Ito, T. (2015). Functional organization of the mammalian auditory midbrain. *J. Physiol. Sci.* 65, 499–506. doi: 10.1007/s12576-015-0394-3

Paxinos, G., Furlong, T., and Watson, C. (2020). *Human Brainstem*, 1st Edn. London: Elsevier Academic Press.

Paxinos, G., Xu-Feng, H., Sengul, G., and Watson, C. (2012). "Organization of brainstem nuclei," in *The Human Nervous System*, eds J. K. Mai and G. Paxinos (Amsterdam: Elsevier), 260–327. doi: 10.1016/b978-0-12-374236-0.10008-2

Powell, E. W., and Hatton, J. B. (1969). Projections of the inferior colliculus in cat. *J. Comp. Neurol.* 136, 183–192. doi: 10.1002/cne.901360205

Ress, D., and Chandrasekaran, B. (2013). Tonotopic organization in the depth of human inferior colliculus. *Front. Hum. Neurosci.* 7:586. doi: 10.3389/fnhum.2013.00586

Rheault, F., De Benedictis, A., Daducci, A., Maffei, C., Tax, C. M. W., Romascano, D., et al. (2020). Tractostorm: the what, why, and how of tractography dissection reproducibility. *Hum. Brain Mapp.* 41, 1859–1874. doi: 10.1002/hbm.24917

Rushmore, R. J., Wilson-Braun, P., Papadimitriou, G., Ng, I., Rathi, Y., Zhang, F., et al. (2020). 3D exploration of the brainstem in 50-Micron resolution MRI. *Front. Neuroanat.* 14:40. doi: 10.3389/fnana.2020.00040

Schilling, K. G., Nath, V., Hansen, C., Parvathaneni, P., Blaber, J., Gao, Y., et al. (2019). Limits to anatomical accuracy of diffusion tractography using modern approaches. *Neuroimage* 185, 1–11. doi: 10.1016/j.neuroimage.2018.10.029

Setsompop, K., Kimmlingen, R., Eberlein, E., Witzel, T., Cohen-Adad, J., McNab, J. A., et al. (2013). Pushing the limits of in vivo diffusion MRI for the Human Connectome Project. *Neuroimage* 80, 220–233. doi: 10.1016/j.neuroimage.2013.05.078

Sigalovsky, I. S., and Melcher, J. R. (2006). Effects of sound level on fMRI activation in human brainstem, thalamic and cortical centers. *Hear. Res.* 215, 67–76. doi: 10.1016/j.heares.2006.03.002

Sitek, K. R., Gulban, O. F., Calabrese, E., Johnson, G. A., Lage-Castellanos, A., Moerel, M., et al. (2019). Mapping the human subcortical auditory system using histology, postmortem MRI and in vivo MRI at 7T. *Elife* 8:e48932. doi: 10.7554/eLife.48932

Smith, R., Raffelt, D., Tournier, J.-D., and Connelly, A. (2020). Quantitative streamlines tractography: methods and inter-subject normalisation. *Preprint* doi: 10.31219/osf.io/c67kn

Smith, R. E., Tournier, J.-D., Calamante, F., and Connelly, A. (2012). Anatomically-constrained tractography: improved diffusion MRI streamlines tractography through effective use of anatomical information. *Neuroimage* 62, 1924–1938. doi: 10.1016/j.neuroimage.2012.06.005

Smith, R. E., Tournier, J.-D., Calamante, F., and Connelly, A. (2015). SIFT2: enabling dense quantitative assessment of brain white matter connectivity using streamlines tractography. *Neuroimage* 119, 338–351. doi: 10.1016/j.neuroimage.2015.06.092

Sotiropoulos, S. N., Jbabdi, S., Xu, J., Andersson, J. L., Moeller, S., Auerbach, E. J., et al. (2013). Advances in diffusion MRI acquisition and processing in the Human Connectome Project. *Neuroimage* 80, 125–143. doi: 10.1016/j.neuroimage.2013.05.057

Straka, M. M., Hughes, R., Lee, P., and Lim, H. H. (2015). Descending and tonotopic projection patterns from the auditory cortex to the inferior colliculus. *Neuroscience* 300, 325–337. doi: 10.1016/j.neuroscience.2015.05.032

Suga, N. (2020). Plasticity of the adult auditory system based on corticocortical and corticofugal modulations. *Neurosci. Biobehav. Rev.* 113, 461–478. doi: 10.1016/j.neubiorev.2020.03.021

Suga, N., Yan, J., and Zhang, Y. (1997). Cortical maps for hearing and egocentric selection for self-organization. *Trends Cogn. Sci. (Regul. Ed.)* 1, 13–20. doi: 10.1016/S1364-6613(97)01002-4

Tabas, A., Mihai, G., Kiebel, S., Trampel, R., and von Kriegstein, K. (2020). Abstract rules drive adaptation in the subcortical sensory pathway. *Elife* 9:e64501. doi: 10.7554/eLife.64501

Tarabichi, O., Kozin, E. D., Kanumuri, V. V., Barber, S., Ghosh, S., Sitek, K. R., et al. (2018). Diffusion tensor imaging of central auditory pathways in patients with sensorineural hearing loss: a systematic review. *Otolaryngol. Head Neck Surg.* 158, 432–442. doi: 10.1177/0194599817739838

Tendler, B. C., Hanayik, T., Ansorge, O., Bangerter-Christensen, S., Berns, G. S., Bertelsen, M. F., et al. (2021). The Digital Brain Bank: an open access platform for post-mortem datasets. *BioRxiv* [Preprint]. doi: 10.1101/2021.06.21.449154

Thomas, C., Ye, F. Q., Irfanoglu, M. O., Modi, P., Saleem, K. S., Leopold, D. A., et al. (2014). Anatomical accuracy of brain connections derived from diffusion MRI tractography is inherently limited. *Proc. Natl. Acad. Sci. U.S.A.* 111, 16574–16579. doi: 10.1073/pnas.1405672111

Tournier, J.-D., Calamante, F., Gadian, D. G., and Connelly, A. (2004). Direct estimation of the fiber orientation density function from diffusion-weighted MRI data using spherical deconvolution. *Neuroimage* 23, 1176–1185. doi: 10.1016/j.neuroimage.2004.07.037

Vu, A. T., Auerbach, E., Lenglet, C., Moeller, S., Sotiropoulos, S. N., Jbabdi, S., et al. (2015). High resolution whole brain diffusion imaging at 7T for the Human Connectome Project. *Neuroimage* 122, 318–331. doi: 10.1016/j.neuroimage.2015.08.004

Wang, F., Dong, Z., Tian, Q., Liao, C., Fan, Q., Hoge, W. S., et al. (2021). In vivo human whole-brain Connectom diffusion MRI dataset at 760 μm isotropic resolution. *Sci. Data* 8:122. doi: 10.1038/s41597-021-00904-z

Webster, D. B. (1992). "An overview of mammalian auditory pathways with an emphasis on humans," in *The Mammalian Auditory Pathway: Neuroanatomy* Springer Handbook of Auditory Research, eds D. B. Webster, A. N. Popper, and R. R. Fay (New York, NY: Springer New York), 1–22.

Winer, J. A. (2005). Decoding the auditory corticofugal systems. *Hear. Res.* 207, 1–9. doi: 10.1016/j.heares.2005.06.007

Winer, J. A., Larue, D. T., Diehl, J. J., and Hefti, B. J. (1998). Auditory cortical projections to the cat inferior colliculus. *J. Comp. Neurol.* 400, 147–174. doi: 10.1002/(SICI)1096-9861(19981019)400:2<147::AID-CNE1<3.0.CO;2-9

Wu, C. M., Ng, S. H., Wang, J. J., and Liu, T. C. (2009). Diffusion tensor imaging of the subcortical auditory tract in subjects with congenital cochlear nerve deficiency. *AJNR Am. J. Neuroradiol.* 30, 1773–1777. doi: 10.3174/ajnr.A1681

Yeh, C.-H., Jones, D. K., Liang, X., Descoteaux, M., and Connelly, A. (2021). Mapping structural connectivity using diffusion MRI: challenges and opportunities. *J. Magn. Reson. Imaging* 53, 1666–1682. doi: 10.1002/jmri.27188

Yeh, C.-H., Smith, R. E., Dhollander, T., Calamante, F., and Connelly, A. (2019). Connectomes from streamlines tractography: assigning streamlines to brain parcellations is not trivial but highly consequential. *Neuroimage* 199, 160–171. doi: 10.1016/j.neuroimage.2019.05.005

Yeh, F.-C., Wedeen, V. J., and Tseng, W.-Y. I. (2010). Generalized q-sampling imaging. *IEEE Trans. Med. Imaging* 29, 1626–1635. doi: 10.1109/TMI.2010.2045126

Yendiki, A., Aggarwal, M., Axer, M., Howard, A. F., van Cappellen van Walsum, A.-M., and Haber, S. N. (2021). Post mortem mapping of connectional anatomy for the validation of diffusion MRI. *BioRxiv* [Preprint]. doi: 10.1101/2021.04.16.440223

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Sitek, Calabrese, Johnson, Ghosh and Chandrasekaran. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership